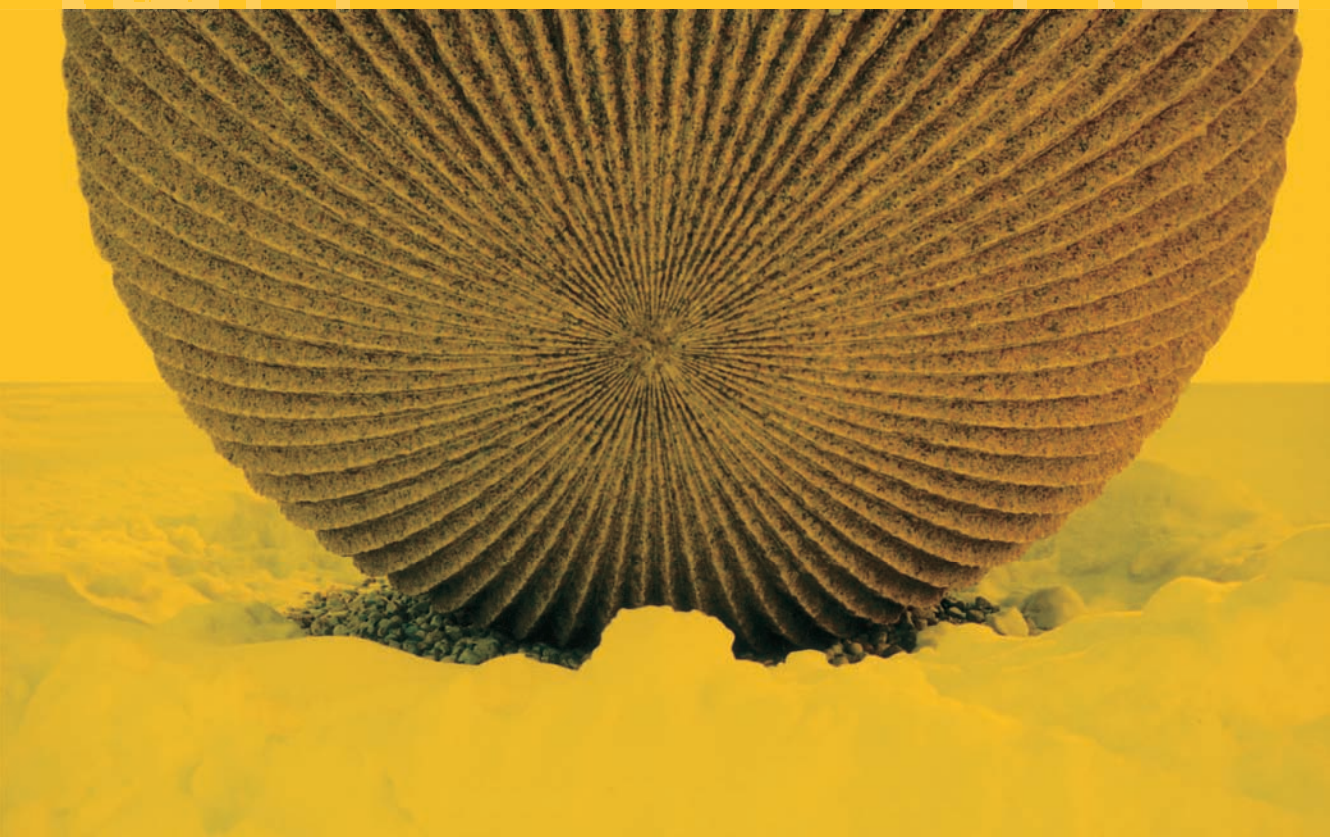
Studia humaniora ouluensia

Kari Suomi, Juhani Toivanen & Riikka Ylitalo

FINNISH SOUND STRUCTURE

Phonetics, phonology, phonotactics and prosody



STUDIA HUMANIORA OULUENSIA 9

KARI SUOMI, JUHANI TOIVANEN & RIIKKA YLITALO

FINNISH SOUND STRUCTURE
Phonetics, phonology, phonotactics and prosody

UNIVERSITY OF OULU, OULU, FINLAND 2008

Studia humaniora ouluensia 9

Publishing office and distribution:
Faculty of Humanities
Linnanmaa
P.O.Box 1000
90014 University of Oulu
Finland

ISBN 978-951-42-8983-5
ISSN 1796-4725 (URL: http://herkules.oulu.fi/issn17964725)

ALSO AVAILABLE IN ELECTRONIC FORMAT
ISBN 978-951-42-8984-2 (URL: http://herkules.oulu.fi/isbn9789514289842)

Typesetting: Heikki Keränen
Cover Design: Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2008

Preface

This book attempts at a description of the sound structure of Standard Spoken Finnish, intended for an international audience familiar with the basic concepts of phonetics, phonology and linguistics. No prior knowledge of Finnish, a Finno-Ugric (and ultimately Uralic) language, is presupposed. The book describes the phonemes and their allophones, the phonotactics and the prosodic system of the language, and it is based on the corresponding parts of our textbook in Finnish (Suomi, Toivanen & Ylitalo 2006), albeit updated and adapted in many ways to the intended readership. To our knowledge, no equally comprehensive description of Finnish sound structure is currently available. The description of the prosodic system is to a considerable extent based on our own recent research.

Including phonotactics along with segmental phonetics and prosody may seem an odd decision. However, we feel that the inclusion of phonotactics is warranted for at least four reasons. Firstly, Finnish is a full-fledged quantity language in which both consonant and vowel durations are contrastive, independently of each other, and of word stress, and according to the standard phonological interpretation, the quantity opposition is a matter of phonotactics: in a given word position, there may be a contrast between one phoneme or a sequence of two identical phonemes. Secondly, given the standard interpretation of the quantity opposition, sequences of up to four vowel phonemes in a word are possible; across a word boundary, even longer sequences of vowels can occur. Thirdly, Finnish has vowel harmony, as a result of which only certain vowels can co-occur in a word; in our view, vowel harmony is best described as a phonotactic restriction, although it is sometimes treated as a prosodic property. Fourthly, in many descriptions of Finnish available in English, the phonotactics of especially word-initial consonants is described in a way that is clearly unrealistic in view of the situation in modern Standard Spoken Finnish. We therefore feel that excluding these phonotactic properties would result in an inadequate picture of the sound structure of the language.

The book aims at a description of the Finnish sound structure, and the descriptive frameworks used in the various chapters of the book are mostly intentionally as theoretically shallow and neutral as has seemed possible. The intention is to provide primary data on Finnish, data that researchers with different theoretical inclinations hopefully find useful for their own purposes.

Segment durations serve both lexical and postlexical functions in Finnish, in contrast to many other languages, and we therefore often repeat rather detailed numerical results of the original papers. In doing this, we only report results whose statistical significance has been tested in the original papers. It has been our goal that the reader should get an adequate idea of the factors influencing segment durations without necessarily consulting the original papers.

Although morphophonological alternations of many kinds are a characteristic property of Finnish, we do not describe them systematically, but offer a glimpse at them in Chapter 1. For example, Grade Alternation is a common type of morphophonological alternations in Finnish. The nominative singular form of 'lamb' is *lammas,* the genitive singular form is *lampaan.* That is, the weak grade /mm/ sequence alternates with the strong grade sequence /mp/. We do not describe such alternations systematically for two reasons. Firstly, morphophonological alternations are properties of morphemes and, phonologically and phonetically, segments that participate in alternations in their respective morphemes do not differ from those segments that do not participate in any alternation in their respective morphemes; at least, there is no phonological difference in models of phonology that are relatively surface-oriented. Thus e.g. the sequence /mm/ in *lammas* is not, at least phonetically, in any way different from the same sequence in *tamma* 'mare', a word that does not participate in grade alternation. Secondly, there are descriptions of Finnish morphophonological alternations available in English, albeit not exhaustive ones, in grammars such as Karlsson (1999) and Sulkala & Karjalainen (1992); readers interested in aspects of Finnish not dealt with in this book are advised to consult these sources.

Besides offering a glimpse at morphophonological alternations, Chapter 1 also briefly exemplifies inflection and word formation by derivation; these often involve morphophonological alternations. Hopefully, Chapter 1 gives an adequate overall picture of the structure of Finnish words.

We wish to thank Matthew Gordon (University of California, Santa Barbara) for useful comments on the manuscript of this book. Any errors and inconsistencies that remain are our own.

Oulu, December 2008

Kari Suomi Juhani Toivanen Riikka Ylitalo

Contents

1. Standard Spoken Finnish: definition and some structural properties

The book aims at describing the sound structure of Standard Spoken Finnish (SSF). This is a form of speech that is used in the educational system and in the media across the country. Originally, it was based on Standard Written Finnish, which in turn was consciously created, in the nineteenth century, as a compromise between the various dialects. In contrast to many other standard languages, then, Standard Finnish (written or spoken) is not based on the language spoken in the centre of power. Consequently, at first, SSF was nobody's native dialect, and it was spoken, in formal situations, by a small number of educated people only; most of the educated and other upper class people at those times spoke Swedish as their native language. Later, SSF has been actively and successfully propagated through the educational system.

Today, SSF may be the native dialect for a number of speakers, but most people learn a local dialect first, and then SSF. Most speakers today have command over two varieties of spoken Finnish: their local dialect and SSF. Usually, the former is used in informal speaking situations, the latter in formal ones; however, some speakers, especially elderly ones, do not necessarily speak SSF on any occasion — and even many younger people never have the chance or duty to speak in formal situations. Although SSF is spoken across the country, it is not a monolith: it has local colourings, especially as concerns prosody, notably segment durations in certain word positions and the way sentence accents are realised; this is clear from so far unpublished results of the third author. It is not expected in the Finnish society that an educated speaker should, in a formal speaking situation, speak SSF according to the strictest norms (recommended if not demanded by an advisory board funded by the state); instead, local colourings are both used and tolerated. That is, local varieties of SSF do not stigmatise the speaker as they do in many other countries, language areas and cultures; recall that we are talking about local varieties of SSF, and not about local vernaculars. Moreover, there is increasing variability in SSF due to register, so that increasingly informal forms of SSF are emerging, and their use in formal speaking situations is increasing; these informal forms include features such as deletions of certain segments in certain positions, or replacements of phonemes of foreign origin by fully native ones, as will be explained in more detail below. In fact, speech according to the strictest norms is nowadays used in highly formal

situations only, and not always so. Speeches by statesmen, and interviews between reporters and ordinary folk (including children), recorded a few decades ago, often sound ridiculous because of their excessive formality, which is very much signalled by the prosody. Speakers also differ among themselves with respect to the conditions of formality under which they switch from the vernacular to SSF and vice versa. For the colloquial spoken language, see also Chapter 22 in Karlsson (1999). While we aim at describing only the sound structure of SSF systematically, we make occasional references to differences between SSF and local dialects; we would gladly make more systematic references to this effect if there were more reliable data available on such differences. When we write that such and such circumstances obtain in Finnish, this means that, as far as we know, the circumstances obtain for Finnish in general.

Morphophonological alternations are very common in Finnish, to the great delight of adult learners of Finnish as a second language. Here only a few examples are given. For now the reader should just take it that all forms written differently are pronounced differently, and that double letters stand for phonetically long segments that contrast with single letters that stand for phonetically short segments. The word *talo* 'house' exhibits no morphophonological alternation, while *susi* 'wolf' does. The forms just given are uninflected, Singular Nominative forms in which both Case and Number always have zero expression. That is, both word forms consist of a stem only. Below these and some inflected Case forms of these words are given, in both Singular (Sg) and Plural (Pl):

	Nominative	Genitive	Essive	Partitive	Illative
Sg	*talo*	*talon*	*talona*	*taloa*	*taloon*
Pl	*talot*	*talojen*	*taloina*	*taloja*	*taloihin*
Sg	*susi*	*suden*	*sutena*	*sutta*	*suteen*
Pl	*sudet*	*susien*	*susina*	*susia*	*susiin*

No attempt has been made here at a morphological analysis of the word forms. In many forms the morphological structure is fully transparent (e.g. *talo+n, talo+na, talo+t*), in others it is more or less opaque and portmanteau morphs exist (e.g. in

susi+en, in which the suffix *en* signals both Plural and Genitive). Notice in passing that Plural is signalled by *t*, *i* or *j* (in forms in which the plural marker is transparent), and that in the inflectional paradigm of *susi*, there is alternation between *s*, *t* and *d*, and between *i*, *e* and *Ø* (zero). The words *talo* and *susi* represent end points on a continuum from no to excessive morphophonological alternation. The alternations in the latter word are relics of past sound chances that were contextually conditioned.

Perhaps the most pervasive morphophonological alternation in Finnish is Grade Alternation that affects consonants. There is both quantitative and qualitative grade alternation. In quantitative grade alternation *p* alternates with *pp*, *t* with *tt*, and *k* with *kk*. Below are examples in two singular cases.

Strong grade		Weak grade	
lippu 'flag'	(Nominative)	*lipun*	(Genitive)
katto 'roof'	(Nominative)	*katon*	(Genitive)
neilikka 'carnation'	(Nominative)	*neilikan*	(Genitive)
ryppään 'cluster'	(Genitive)	*rypäs*	(Nominative)
rattaan 'wheel'	(Genitive)	*ratas*	(Nominative)
rakkaan 'dear'	(Genitive)	*rakas*	(Nominative)

In the first three examples, the strong grade appears in the singular nominative case (as it does in some other cases), the weak grade in the singular genitive case (as it does in some other cases). The last three examples exhibit reverse grade alternation: the grades occur in the opposite sets of cases.

In qualitative grade alternation, qualitatively different consonants (or consonant sequences) alternate with each other or with zero, e.g. *mp* with *mm*, *k* with *Ø*, *t* with *d*. Some examples, again in two cases:

Strong grade		Weak grade	
kampa 'comb'	(Nominative)	*kamman*	(Genitive)
parta 'beard'	(Nominative)	*parran*	(Genitive)
luku 'number'	(Nominative)	*luvun*	(Genitive)
laki 'law'	(Nominative)	*lain*	(Genitive)
lampaan 'lamb'	(Genitive)	*lammas*	(Nominative)
mateen 'burbot'	(Genitive)	*made*	(Nominative)
varpaan 'toe'	(Genitive)	*varvas*	(Nominative)
aikeen 'intention'	(Genitive)	*aie*	(Nominative)

These are just examples, further alternating patterns exist. The last four examples again exhibit reverse alternation.

Finnish is a heavily inflected language. Instead of mainly using prepositions and word collocations, many kinds of suffixes are added to word stems (but also prepositions as well as postpositions exist). Nouns, adjectives and nominalised forms of verbs usually have 15 Cases, both in singular and plural (but as already mentioned, both Case and Number always have zero expression in the singular nominative forms). In order not to make things too difficult, the word *talo*, with no stem-internal morphophonological alternation, has been chosen as the example of nominal inflection, with morpheme boundaries indicated; some forms of the first person pronoun are also given to represent the richer paradigm of the pronouns:

	Singular	Plural
Nominative	*talo, minä*	*talo+t, me*
Genitive	*talo+n, minun*	*talo+je+n, meidän*

Partitive	*talo+a*	*talo+j+a*
Accusative	*talo, talo+n, minut*	*talo+t, meidät*
Essive	*talo+na*	*talo+i+na*
Translative	*talo+ksi*	*talo+i+ksi*
Inessive	*talo+ssa*	*talo+i+ssa*
Elative	*talo+sta*	*talo+i+sta*
Illative	*talo+on*	*talo+i+hin*
Adessive	*talo+lla*	*talo+i+lla*
Ablative	*talo+lta*	*talo+i+lta*
Allative	*talo+lle*	*talo+i+lle*
Comitative	*talo+ine+*	*talo+i+ne+*
Abessive	*talo+tta*	*talo+i+tta*
Instructive	--	*talo+in*

Notice that the plural marker follows immediately after the stem, and before the case endings. For almost all nominally inflected words, Accusative has two forms in Singular, one identical to the Nominative form, the other identical to the Genitive form; in Plural, the Accusative form is identical to the Nominative form. But pronouns have separate forms even for each of Nominative, Accusative and Genitive as in the examples *minä* 'I', *me* 'we', *minun* 'my', *meidän* 'our', etc. Comitative forms must always be followed by a possessive suffix, e.g. *talo+ine+ni* 'with my house(s)', *talo+ine+mme* 'with our house(s)'; Comitative forms are ambiguous as to Number. The Instructive case only exists in Plural, *talo+in* means 'with the help of houses'. The last three cases are not as commonly used as the other ones; the Abessive form *talo+tta* means 'without a/the house' (Finnish uses means other than articles to denote definiteness).

Case suffixes are often followed by further suffixes, for example *talo+i+ssa+ni+kin* 'in my houses, too' (*talo* 'house' + *i* 'many' + *ssa* 'in' + *ni*

‘my’ + *kin* ‘also’). The suffix *-kin* is a clitic, and there are five others, with pragmatic meanings. If all these are maximally utilised, a noun can have 2253 different word forms: to see all these for *kauppa* ‘shop’ go to Fred Karlsson’s website at: <http://www.ling.helsinki.fi/~fkarlsso/genkau2.html>. All these forms are grammatical, but many of them are rather contrived and would be seldom used because of the many simultaneous pragmatic meanings.

Verbs express Tense, Mood and Person by inflectional suffixes; however, both Indicative and Present Tense have zero expression:

1. Person	*laula+n*	’I sing’	*laula+mme*	’we sing’
2. Person	*laula+t*	’thou singst’	*laula+tte*	’you sing’
3. Person	*laula+a*	‘(s)he sings’	*laula+vat*	’they sing’
4. Person	*laule+taan*	’somebody sings’		

Indicative Imperfect (’I sang’, ‘thou sangst’, ‘(s)he sang’, etc):

1. Person	*laulo+i+n*	*laulo+i+mme*
2. Person	*laulo+i+t*	*laulo+i+tte*
3. Person	*laulo+i*	*laulo+i+vat*
4. Person	*laule+tt+iin*	

Indicative Perfect and Indicative Plusquamperfect use the same suffixes, and differ from each other with respect to the form of the auxiliary verb (the verb ‘to be’). So Indicative Perfect has the following forms, in which the auxiliary verb has Present tense (’I have sung’, ’thou hast sung’, ’(s)he has sung’ etc.):

1. Person	*olen laula+nut*	*olemme laula+neet*
2. Person	*olet laula+nut*	*olette laula+neet*
3. Person	*on laula+nut*	*ovat laula+neet*

4. Person	*on laule+ttu*	

Indicative Plusquamperfect has the same inflectional endings as Indicative Perfect, but the auxiliary verb has Past tense ('I had sung' etc.):

1. Person	*olin laula+nut*	*olimme laula+neet*
2. Person	*olit laula+nut*	*olitte laula+neet*
3. Person	*oli laula+nut*	*olivat laula+neet*
4. Person	*oli laule+ttu*	

Another Mood in addition to Indicative is Conditional ('I would sing', 'thou wouldst sing', '(s)he would sing' etc.):

1. Person	*laula+isi+n*	*laula+isi+mme*
2. Person	*laula+isi+t*	*laula+isi+tte*
3. Person	*laula+isi*	*laula+isi+vat*
4. Person	*laule+tta+isi+in*	

Potential Present ('I may sing' etc):

1. Person	*laula+ne+n*	*laula+ne+mme*
2. Person	*laula+ne+t*	*laula+ne+tte*
3. Person	*laula+ne+e*	*laula+ne+vat*
4. Person	*laule+tta+ne+en*	

Imperative ('sing!', 'let him/her sing', 'let us sing', etc.):

1. Person	--	*laula+kaa+mme*
2. Person	*laula*	*laula+kaa*
3. Person	*laula+koon*	*laula+koot*
4. Person	*laule+tta+koon*	

Altogether, a normally inflected verb has 528 finite forms, of which some examples were given above. In addition, a normally inflected verb has 324 infinitive forms and about 11000 participial forms that are inflected like nouns, thus altogether about 12 000 forms (grammatical and phonological words, consisting of stem + suffixes), see Karlsson (1983: 356–357). These forms do not include derivative suffixes. If these were included, the numbers would become manifold.

To conclude this structural sketch, let us look at some examples of word formation by derivation. Words are extensively formed by adding one or more derivative suffixes to a stem, often with accompanying morphophonological alternation. Examples (in some of the example words, the morphological analysis could go further, and the glosses given are not the only ones):

järki	'sense'
järje+tön	'senseless'
järje+stö	'organisation'
järje+stä+ä	'to organise'
järje+stä+yty+mi+nen	'formation'
järje+stelmä	'system'
järje+stelmä+lli+nen	'systematic'
järje+stelmä+lli+syys	'habit of being systematic'

The fourth example is the 1^{st} infinitive of a verb, the fifth one is a nominal form of a verb, the rest are derived nouns and adjectives; all word forms except the infinitive are nominal words in nominative singular, and can be inflected in all nominal cases. The 14-syllable example word below, a noun in singular adessive case, is somewhat far-fetched, but it does demonstrate the possibilities of word formation by derivation:

järjestelmällistyttämättömyydellänsäkään 'not even with his/her lack of systematization'.

Roughly, the morphemes are as follows: järje+stelmä+llis 'systemati' + tyttä 'ization' + mättö+myyde 'lack of' + llä 'with' + nsä 'his/her' + kään 'not even'. Despite its length, this is clearly one word.

2 Airstream mechanisms and phonation

For most of the time, Finnish is spoken using the pulmonic egressive airstream. However, the pulmonic ingressive airstream is occasionally used by some speakers in very short utterances, e.g. *joo, juu, nii(n)* (colloquial for 'yes'), or at the end of longer, otherwise egressively produced utterances. Moreover, some speakers occasionally produce considerable stretches of speech ingressively. It is difficult to assign any specific meaning (pragmatic or otherwise) to this manner of speaking. The ingressiveness is often noticed and commented on by speakers of languages in which ingressive speech presumably does not occur, or at least not as often. Non-pulmonic airstream mechanisms are not used.

Differences in the mode of phonation are not used in Finnish for directly linguistic purposes, e.g. to distinguish lexical meanings. Ladefoged & Maddieson (1996: 49) recognise five steps in the continuum of modes of vibration of the glottis, namely breathy voice, slack voice, modal voice, stiff voice and creaky voice. Here, when we refer to deviations from modal voice, we only use the terms breathy voice and creaky voice, which in the classification by Ladefoged and Maddieson represent the end points of the continuum. However, in using these terms, we do not and cannot take any stance on the differences between the end points and the intermediate non-modal steps in the continuum. Thus in a more detailed and expert analysis, instances of what we call breathy voice might be characterised as slack voice, and similarly with creaky and stiff voice. Consequently, breathy voice here simply means a perceptible deviation from modal voice in one direction, and creaky voice a perceptible deviation in the other direction. Modal phonation is used by most speakers most of the time, and deviations from modal phonation that characterise speech as a whole are usually speaker-specific properties or accompany certain emotional states (but for creaky phonation by young females, see immediately below). However, creaky and breathy voice, as well as whisper, are quite common near the ends of utterances, as will be discussed in Chapter 10 and summarised in section 10.10.

Creaky phonation, as an overall property of speech irrespective of prosodic boundaries, is clearly becoming an increasingly common (and to many, objectionable) hallmark of young, especially female speakers. It is not easy to pinpoint the conditions of this fashionable overall creak, but it seems to have social underpinnings. It may have no other function than to indicate some kind of possibly fully unconscious affiliation with fashionable groups of speakers. It

seems to us, for example, that a woman working in a factory is less likely to use overall creak than a woman studying literature at the university. We also have a feeling that the creak is more common in the south, near the centres of power, than in the national periphery, but this feeling may be due simply to the fact that people from the periphery get their voices heard less often than those living in the centres. Anyway, it seems to us that a redneck male farmer would not use the creak, unless it was his physiologically determined manner of phonation.

3 Phonemes and allophones

In the phonemic analysis adopted here, contrastively long segments are interpreted as sequences of two identical phonemes. This is the standard interpretation, and it will be motivated in Chapter 4 below, at a point at which all circumstances relevant to the interpretation have been presented. In the orthography, any sequence of two identical phonemes is represented by a sequence of two identical graphemes, and there are words like *ta.ka* [t̪ɑkɑˑ] /t̪ɑkɑ/, *taa.ka* [t̪ɑːkɑ̆] /t̪ɑɑkɑ/, *tak.ka* [t̪ɑkːɑ̆] /t̪ɑkkɑ/, *taak.ka* [t̪ɑːkːɑ̆] /t̪ɑɑkkɑ/, *ta.kaa* [t̪ɑkɑː] /t̪ɑkɑɑ/, *taa.kaa* [t̪ɑːkɑː] /t̪ɑɑkɑɑ/, *taak.kaa* [t̪ɑːkːɑː] /t̪ɑɑkkɑɑ/ (with syllable boundaries indicated in the orthographic forms). Details of vowel and consonant durations will be discussed in section 9.3 below. Diphthongs, sequences of two dissimilar vowels, are structurally equivalent to sequences of two identical vowel phonemes, e.g. *tai.ka* is phonemically /t̪ɑikɑ/ in which /ɑ/ and /i/ are separate phonemes in the first syllable. Qualitatively, phonetically short and long (phonemically single and double) vowels sound very similar to a native speaker's ears. Wiik (1965: 56–60) reported slightly more centralised F1 and F2 frequencies for single than for double vowels in the primarily-stressed syllable, the numerical differences being greatest in the high vowels but very small in the low vowels. Unfortunately, Wiik performed no statistical analyses, and the identities of the abutting consonants were not controlled.

However, O'Dell (2003: 73–74) extracted spectral, intensity and fundamental frequency differences from natural accented tokens of *tuli* and *tuuli* produced in a constant frame sentence, and synthesised, using dynamic time warping, two series of eleven stimuli that preserved these differences. In both series, segment durations were stepwise varied so that the extreme stimuli, the first one and the eleventh one, fully corresponded to the original words. That is, one series contained the qualitative parameters calculated from original *tuli*, the other series used the qualitative parameters from original *tuuli*. Listeners were asked to categorise the stimuli in both series as either *tuli* or *tuuli*, and their responses were clearly affected by the qualitative differences except at the extreme ends of the two series. Inspection of the spectral differences indicated that the single /u/ in *tuli* had higher F1 and F2 frequencies than the double /uu/ in *tuuli*, i.e. that the single vowel was centralised relative to the double one. Interestingly, an opposite effect was observed in the word-final /i/: this vowel was more centralised in *tuuli* than in *tuli*. As will be shown below, the second-syllable /i/ in a CVCV word like

tuli has about twice the duration of the same phoneme in a CVVCV word like *tuuli*, and the phonetically short vowel was centralised relative to the phonetically long one. O'Dell's results are thus consistent with those of Wiik (1965) in that there is some centralisation in the high single vowels relative to high double vowels. But O'Dell's results suggest that the centralisation is determined by the duration and not by the phonological quantity of a vowel (as indicated by the second-syllable /i/ vowels). This sounds plausible, and O'Dell's observations question earlier claims that spectral differences between single and double vowels have no effect on quantity perception; see e.g. Lehtonen (1970: 21–22; 87). But the effects may be restricted to non-low vowels.

The allophones identified below have been arrived at by a number of different means: auditory and acoustic analyses as well as articulatory introspection.

3.1 Vowels

Using the nearest IPA cardinal vowel symbols, the eight vowel phonemes could be given as /i/, /e/, /y/, /ø/, /æ/, /ɑ/, /o/ and /u/ (but on the notations /e/, /ø/ and /o/ see below). They occur e.g. in the series of word forms *mikin – mekin – mykin – mökin – mäkin – makin – mokin – mukin*. The form *mikin* is the genitive singular form of *Mikki* (and of colloquial *mikki* 'microphone), *mekin* is *me* 'we' (nominative) + *kin* 'also' = 'we, too', *mykin* is the plural instructive of *mykkä* 'dumb', *mökin* is the singular genitive of *mökki* 'cottage', *mäkin* is the singular genitive of *Mäkki* (colloquial for *MacIntosh*) as well as *mä* (colloquial for *minä* 'I') + *kin* 'also', *makin* is the genitive singular of *maki* 'lemur', *mokin* is the plural instructive of *mokka* 'suede' and of *moka* 'mistake', and *mukin* is the singular genitive of *muki* 'mug'.

A given vowel phoneme is always written with the same grapheme. This is true of most consonant phonemes, too, and below, example words are often given in their phonemically unambiguous orthographic forms. Diphthongs are analysed as combinations of the eight vowel phonemes. The greatest discrepancy between the IPA vowels and the Finnish ones concern the Finnish mid series /e/, /ø/, and /o/. These vowels are approximately half-way between the IPA [e] and [ɛ], [ø] and [œ], and [o] and [ɔ], respectively, and they could thus be transcribed as either /e̞/, /ø̞/, /o̞/, or as /ɛ̝/, /œ̝/, /ɔ̝/, respectively. Otherwise, the Finnish vowels are somewhat less extreme than the respective nearest IPA vowels. In particular, /ɑ/

has the same height as /æ/; Maddieson (1984) classifies Finnish /ɑ/ as a low central vowel. For acoustic comparisons of Finnish vowels with English, French and Swedish vowels see Wiik (1965), Vihanta (1978) and Kuronen (2000), respectively. Figure 1 is a schematic summary of the results of F1 and F2 measurements in monophthongs.

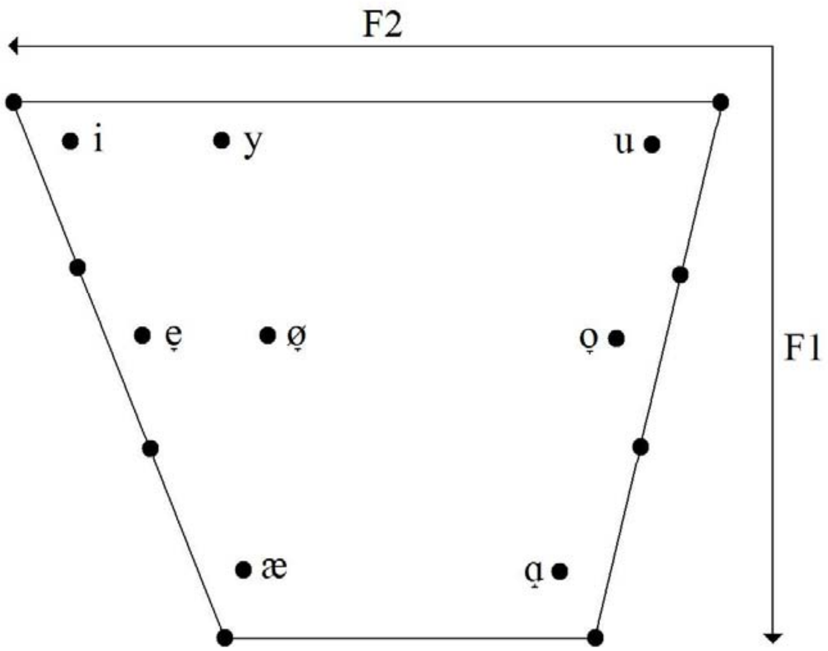


Fig. 1. The approximate locations of the Finnish vowel phonemes in a two-dimensional F1 – F2 vowel space.

In the following, for simplicity, the IPA cardinal vowel symbols without diacritics will be used. Finnish thus has five peripheral vowels of the sort that are typologically common (/i/, /e/, /ɑ/, /o/, /u/), and in addition three vowels that are typologically less common (/y/, /ø/, /æ/). As will be explained in more detail in section 6.1.2 below, /ɑ/, /o/ and /u/ act as a class in vowel harmony, as do /y/, /ø/ and /æ/, and /y/ alternates with /u/, /ø/ with /o/, and /ɑ/ with /æ/ in suffixal vowel harmony. According to both formant measurements and perceptual criteria, the

vowels clearly group themselves into three height classes: the high vowels /i/, /y/ and /u/, the mid vowels /e/, /ø/ and /o/, and the low vowels /ɑ/ and /æ/.

Compared to the vowel system of e.g. English, the Finnish system can be characterised as symmetrical or neat, as there are no vowels not assignable phonetically to a class consisting of at least two vowels. Moreover, as will be shown below, all of the eight single vowels can also occur double. In British English (in RP at least), for example, /æ/ (as in *hat*) is a short (or lax) vowel on account of its phonotactic behaviour (there cannot be words of the structure */Cæ/, just as there cannot be words of the structure */Cɪ/, /ɪ/ being the vowel that occurs in e.g. *hit*), but on account of its duration, /æ/ behaves like the long (or tense) vowels. And what would be the long counterpart of /æ/ in the English vowel system?

Finnish vowels undergo nasalisation in the vicinity of nasal consonants, especially between tautosyllabic nasals. Thus the first-syllable vowel in *mamma* '(grand)mother' is regularly nasalised, while that in *pappa* '(grand)father' is not; this has been verified in a number of informal comparisons using cross-splicing. Otherwise, the vowels exhibit no noticeable qualitative allophonic variation in addition to that caused by coarticulation; durational alternations will be discussed in section 9.3 below. Stress has at most very little effect on vowel quality — there is no or at most little reduction in unstressed positions — and speech tempo similarly has at most little effect. To our knowledge, potential effects of the sort just mentioned have not been experimentally investigated, and the lack of such experiments suggests, indirectly at least, that such effects must be negligible, if they exist. Especially as concerns the effect of stress on vowel quality, Finnish clearly differs from at least many Germanic languages and Russian. This may have to do with the difference in the stress system: fixed in Finnish, moving in Germanic languages and in Russian. That is, when the location of primary stress is not fully predictable as in these latter languages, it may be necessary to reduce the quality of unstressed so as to make the stressed syllable more salient. However, this is only speculation, and large-scale typological studies would be necessary to assess whether there is any correlation between type of stress system and vowel reduction in unstressed syllables.

In contrast to the consonant system (see below), the vowel system can be characterised as stable. No tendencies towards changes in the system are discernible, such as e.g. new phonemes emerging due to the influence of foreign languages. The vowel /ø/ is the latest newcomer in the language, it is relatively

rare in the vocabulary, and words containing it often have affective and negative connotations, e.g. *hökkeli* 'shack', *törppö* 'scatterbrain', *töllöttää* 'to gape at'. The vowel system is also stable across the dialects, with all dialects having the same eight phonemes. However, the precise phonetic values of at least /ɑ/, /æ/ and /o/ vary somewhat across dialects; for the dialect of Tampere, see Kuronen (2000).

Tautosyllabic vocalic portions can be classified into three groups: single vowels (e.g. /u/ in the first syllable of *tuli* 'fire'), double vowels (e.g. /uu/ in *tuuli* 'wind'), and diphthongs (e.g. /uo/ in *tuoli* 'chair'). In addition there are vowel sequences across a syllable boundary (e.g. /eo/ in *te.os* 'work'). There are also sequences of three and four vowels, but at most two of the vowels can be tautosyllabic, e.g. *ai.e* 'intention', *kaa.os* 'chaos', *ai.oin* 'I intended'. There are good structural grounds for regarding diphthongs as sequences of two phonemes. One of these grounds is the fact that the first and second components of diphthongs all also occur as single vowels. That is, in addition to e.g. the diphthong /ɑi/, also /ɑ/ and /i/ exist, and similarly for the other diphthongs. This in contrast with e.g. the English diphthongs /aɪ/ and /aʊ/ (as in *my* and *now),* in which the first component does not occur as a separate phoneme.

Phonologically, then, diphthongs are sequences of two dissimilar single vowel phonemes. Durationally and metrically (see below), diphthongs are equivalent to double vowels. Acoustically, however, the formant values at the end of a diphthong usually do not reach the target values corresponding to those of the single vowel formally constituting the second member of the diphthong; the tail does not reach the target values corresponding to the phonological analysis. E.g., in a diphthong like /ɑi/ as in *kaide*, the formant values at the end are usually somewhat centralised relative to those of monophthongal /i/ in *kide*. Perceptually and intuitively, however, the diphthong /ɑi/ is a sequence of /ɑ/ and /i/.

3.2 Consonants

In the classification of consonants according to manner of articulation, three major classes are distinguished in this book, namely obstruents, glottals and resonants. The class of glottals is very small, consisting of only [h], [ɦ] and [ʔ] (of which the latter only occurs as an aphonematic segment). The IPA Chart does not recognise (at least up to the version revised to 2005) a separate manner class of glottals (although it recognises glottal as a place of articulation), and [h] and [ɦ] are classified as fricatives, and [ʔ] as a plosive. But the IPA classification is far

from unproblematic. Thus true fricatives can be voiced, voiceless unaspirated or voiceless aspirated as in Burmese (Ladefoged & Maddieson 1996: 179), whereas [h] is voiceless and [ɦ] is breathy voiced. A distinction between aspirated and non-aspirated [h] is an impossibility, and [ɦ] cannot be produced with modal phonation (as can true fricatives). A ternary VOT (Voice Onset Time) opposition is more common in true plosives (as in Thai) than in true fricatives, but [ʔ] is by necessity always voiceless (as indicated in the IPA Chart). Moreover, unlike true fricatives and plosives, [h], [ɦ] and [ʔ] lack a supraglottal place of articulation. In the framework of the IPA classification, then, one would be forced to say something like the following: Fricatives may be distinguished by different values of VOT in the modal register (exception: in fricatives formed at the glottis, only voiceless and breathy voiced fricatives are possible), and plosives may be distinguished by different values of VOT in the modal register (exception: in plosives formed at the glottis, only voiceless ones are possible). To get rid of such exceptions within the class of obstruents, the small class of glottals has been set up in this book. Within this manner class, all constraints on (the mode of) phonation are directly relatable to the circumstances obtaining at the place of articulation: in fact, place and manner are inextricably interwoven in glottals.

It is not possible to state, without qualifications, the number of consonant phonemes in Finnish, as it is to state that of vowel phonemes. It is not possible simply to state that there are X consonant phonemes. The reason for this is that the size of the consonant paradigm is different in different varieties of the language. Due to the differences between the varieties, there are many consonant paradigms. In Karlsson's (1983) terms, Finnish is polysystemic with respect to its consonants.

The Finnish consonant phonemes can be divided into five groups on the basis of how they occur in the different paradigms, i.e. on the basis of which paradigms they belong to or do not belong to. The grouping, shown in Figure 2, has first been suggested by Karlsson (1983).

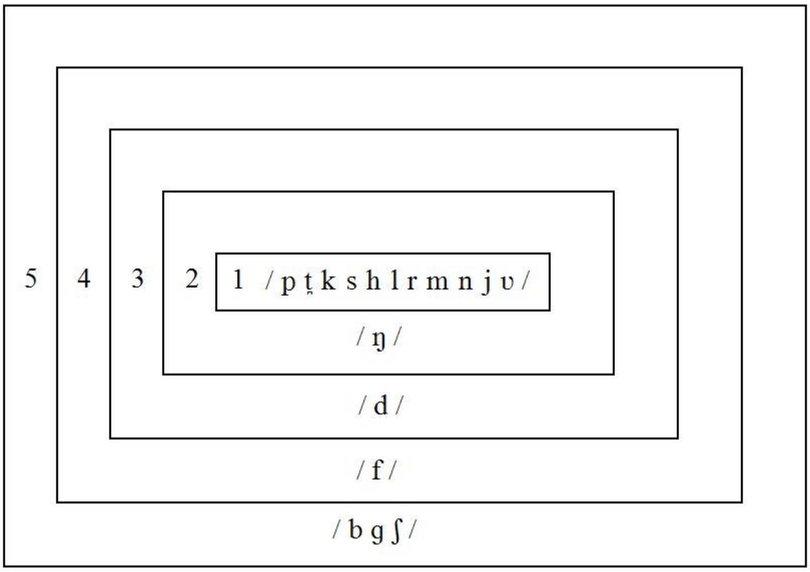


Fig. 2. The groups of Finnish consonant phonemes. Consonant paradigms consist of combinations of the groups.

Group (1) is common to all synchronic varieties, but the other groups belong to only some variants. Group (5) in turn is the most marginal one, i.e. its consonants belong to only some speakers' paradigm, and even for these speakers, not necessarily in all speaking situations. Group (4) belongs to the paradigm of more speakers than group (5) does, group (3) in turn occurs more frequently than group (4), etc. In other words, the larger the group number, the more marginal the group. The minimum consonant paradigm contains only group (1), or 11 phonemes, the maximum consonant paradigm contains groups (1) – (5), or 17 phonemes.

It is often the case that if a paradigm contains a certain group, it also contains all of the lower-numbered groups. For example, if a speaker's paradigm, in a given speaking situation, contains group (5), it is highly probable that the paradigm also contains — in addition to group 1 — groups (2), (3) and (4). However, such inclusiveness does not always hold. There are also varieties that contain e.g. groups (1), (3) and (4), but not group (2); such varieties do contain

[ŋ], but as an allophone of /n/. The conditions under which particular groups belong to paradigms will be discussed below.

Next, the Finnish consonant phonemes and their allophones are described according to the grouping in Figure 2, starting with group (1). Before that, however, a general allophonic statement can be made. For each unrounded allophone (occurring near unrounded vowels, e.g. [l] in *liima* 'glue'), an otherwise identical but rounded allophone also exists (near rounded vowels, e.g. [lʷ] in *luumu* 'plum'). Given this generalisation, the fully predictable rounded allophones will not be separately mentioned below.

Group (1): {/p/, /t̪/, /k/, /s/, /h/, /m/, /n/, /l/, /r/, /ʋ/, /j/}

§ The plosives /p/, /t̪/, /k/

The plosives are voiceless unaspirated and the explosive burst is weak; Suomi (1980: 99) reported the mean VOT values (consisting of the burst and any aspiration), in the word initial position, of 9 ms for /p/, 11 ms for /t̪/ and 20 ms for /k/, and in word medial position, 11 ms, 16 ms and 25 ms, respectively. However, in Suomi (submitted), the target items included nonsense words like *patna*. The sequence /t̪n/ does not occur in fully native words, but there are loans like *etninen* 'ethnic', *luutnantti* 'lieutenant'. The seven speakers produced this sequence in one of two ways: either with a nasal release of the stop closure (with, of course, no aspiration), or with an oral release, often accompanied by a period of aspiration that was considerably longer than that observed for /t̪/ + vowel sequences. Three talkers produced only aspirated /t/'s, one talker produced five aspirated /t/'s and four nasally released ones, two talkers produced two aspirated /t/'s each (and seven nasally released ones), and one talker produced only nasally released /t/'s. This is clearly a special case and, to our knowledge, the only systematic observation of aspiration in the Finnish plosives. It may be noted in passing that Finnish consonants involving a complete closure are, as far as we know, always released before another consonant, except for nasals before a homorganic consonant. Thus in consonant sequences like /t̪n/, /t̪k/, /mn/ the first consonant is always released before the onset of the occlusion of the second consonant. This is in contrasts to the situation in e.g. English in which the first consonant in corresponding sequences is at least often not released.

The plosives can be partly or fully voiced in fast and careless speech, and occasionally even in reading aloud experimental texts in the laboratory. Although usually laminal dentialveolar, /t̪/ has the (pre)alveolar allophone /t/ after /s/ as in *kaste* 'dew'. As in many other languages, /k/ has the more front allophone [k̟] before front vowels; [k], occurring elsewhere, is the main allophone. In this book the classification of places and manners of articulation by Ladefoged and Maddieson (1996) is followed.

§ The sibilant /s/

It must be made clear at the outset that the most common allophone of Finnish /s/ is less "sharp" than the sibilant denoted by the IPA symbol [s]. With respect to its noise (the location of the greatest energy in the spectrum), and perceptually, this allophone is somewhere between IPA [s] and [ʃ]. But since IPA has no symbol or diacritic to correctly characterise this allophone, the notations /s/ and [s] are used here for lack of more accurate symbols. In many varieties of Finnish /s/ is the only sibilant, and also the only fricative if /h/ is classified as a member of the major class of glottals, distinct from the class of obstruents, as is assumed here.

/s/ → [s] ~ [ʃ] (much variation between speakers)

→ [z] / especially in fast speech between vowels

→ [x] / often in the context: __ [r]

Presumably because /s/ is the only sibilant in most varieties, it has plenty of phonetic space for itself without any danger of perceptual confusion. Whether or not this is the correct explanation, there is nevertheless much variation in how /s/ is realised phonetically, roughly from IPA [s] to almost [ʃ]; all of these variable productions are easily identified as /s/ (in those varieties in which there is no /ʃ/). Like the plosives, /s/ is often voiced, in similar circumstances. The allophone [x] occurs often before [r]. The sequence /sr/ is prohibited word-internally in native words, but it occurs in loanwords like *Israel* and *Osram,* across the boundary between the components of compound words, and across full word boundaries. There is an alternative way of pronouncing the /sr/ sequence, and more detailed

discussion of this sequence, and of the allophone [x] of /s/, will be postponed to the description of the allophones of /r/ below.

§ The phoneme /h/*, a glottal continuant or oral fricative, depending on the allophone*

In Finnish, /h/ occurs in many word positions, and its distribution is wider than that of the corresponding phoneme in e.g. the Germanic languages. In these latter languages it is usual for /h/ to occur only word-initially and foot-initially, as in English *hold* and *behold* (i.e., always syllable-initially), but in Finnish /h/ also occurs syllable-finally, as in e.g. *lah.ja*, *vah.ti*, *vaah.t*o, in addition to the syllable-initial position. In syllable-final positions, however, the allophones are oral fricatives, not glottal continuants.

/h/	→	[ç] / between a high front vowel and a consonant
	→	[x] / between a back vowel and a consonant
	→	[ɦ] / between vowels, especially word-internally
	→	[h] / elsewhere

Example words with [ç] are *vihma, pihvi, lyhty, vihje, vihko;* with [x] *tahma, kahvi, tuhti, kohme, tuhka;* with [ɦ] *vihi, vähä, vaha;* and with [h] *haamu, tähti, lehvä.* The phonetic realisation of /h/ is thus very varied (recalling that there are also as many rounded allophones). The fricative allophones [ç] and [x] only occur word-internally in a syllable-final position, before another consonant, but the glottal allophones also occur word-initially and syllable-initially within a word. A preliminary generalisation is that a fricative allophone occurs before another consonant, and a glottal allophone before a vowel.

§ The nasals /m/ *and* /n/

Languages usually have nasals at roughly the same places of articulation as they have plosives. This is true also in Finnish (but while /t̪/ and /n/ are both broadly

coronal, the main allophone of the former is laminal dentialveolar, the latter apicoalveolar). Thus phonetic [ŋ] occurs in all varieties, it is very narrowly phonemic in most varieties, but not in all, and therefore /ŋ/ on its own constitutes group (2), and it is discussed separately below.

/m/ → [ɱ] / usually before [f]

→ [m] / elsewhere

/n/ → [n̻] / _ [t̻], [t̻] _

→ [n] / elsewhere

The allophone [ɱ] is rare because the sequence /mf/ is rare (it occurs in some loanwords, e.g. *amfetamiini* 'amphetamine', *kamferi* 'camphor'). The sequence /n̻t̻/ is very frequent, /t̻n̻/ does not occur in fully native words but there are loanwords like *luutnantti, Botnia.*

§ The lateral approximant /l/

/l/ → [l̻] / _ [t̻], [t̻] _

→ [l] / elsewhere

The allophone [l̻] is laminal dentialveolar when the laminal dentialveolar /t̻/ immediately follows or precedes; the main allophone [l] in turn is apical alveolar. The sequence /t̻l̻/ does not occur in fully native words, but there are fully pronounceable loanwords like *atlas* and *kotletti* 'cutlet', and the sequence also occurs across a word boundary.

§ *The rhotic* /r/

It is not straightforward to determine the main allophone of /r/. Traditionally, it has been stated that the main allophone is a trill, but recent investigations suggest that the tap pronunciation ([ɾ]) is much more common than has been previously presumed. Thus Mustanoja & O'Dell (2007) observed, in two corpora (colloquial speech in Tampere and news broadcasts of the national broadcasting company) that a great majority of the single /r/ productions in word-medial intervocalic position were taps (75% and 90% in the two corpora, respectively). It seems however that in other positions, e.g. word-initially, a trill realisation is more common than the tap. Also the quantity opposition has an effect on the realisation of /r/. Thus, as just mentioned, a single /r/ (as in *paras* 'best') is often realised as [ɾ], but double /rr/ (as in *parras* 'edge') always as [rː], often with several closure periods. On the whole, it seems legitimate to conclude that the main allophone of /r/ is [r].

In addition, /r/ has an alveolar fricative allophone, [ɹ̝]. As was mentioned above in discussing the allophones of /s/, the sequence /sr/ does not occur inside native uncompounded words. A very likely reason for this avoidance is that the sequence of the main allophones of the two phonemes, i.e. [sr], a sequence consisting of a sibilant and a trill, is difficult to pronounce even for a native phonetician. However, the phoneme sequence /sr/ is common across a word boundary, and across a boundary between the parts of compound words. One way of avoiding the pronunciation difficulty in e.g. *Israel* is that the allophonic sequence [sɹ̝] is chosen. Another way, mentioned earlier, is to choose the allophonic sequence [xr]. Thus, the difficult sequence sibilant + trill is avoided either by replacing the trill by a fricative rhotic, or by replacing the sibilant by a non-sibilant fricative. The sequence [xɹ̝] is not attested. The following account of the allophones of /r/ is somewhat tentative:

/r/	→	[ɹ̝] / [s] _
	→	[ɾ] / word internally
	→	[r] / elsewhere (especially in /rr/)

In addition to these allophones proper, /r/ has more marginal realisations that occur in some speakers' idiolects, realisations that are not dialectal properties nor due to segmental context. These include [ɹ] and [ʁ]. There is very little areal variation in the regular allophones of /r/, but in the Tampere area their place of articulation is further back than in other dialects, probably apical retroflex (Kuronen 2000).

§ The central approximants /ʋ/ and /j/

As in many other languages, the Finnish central approximants (or semivowels) only occur in the syllable onset position.

/j/	→	[j]
/ʋ/	→	[w] / after diphthongs ending in [u]
	→	[ʋ] / elsewhere

The allophone [w] of /ʋ/ occurs in e.g. *sauva* [sɑuwɑ] 'staff' and *rouva* [rouwɑ] 'married woman'.

The consonant phonemes to be discussed below have not been observed to exhibit noteworthy allophonic variation (apart from the rounded-unrounded variation), and hence no allophonic statements are given.

***Group (2)**: {/ŋ/}*

The phonetic nasal [ŋ] occurs in all varieties of Finnish. In the native vocabulary, [ŋ] has a very narrow distribution: phonetically short it occurs only in the word-medial context /_ /k/ (e.g. *lanka* [lɑŋkɑ] 'thread'), and phonetically long only in the context /V_V (*sangen* [sɑŋːen] 'very'). In the context /_ /k/, [ŋ] is in complementary distribution with the nasals [m] and [n] that do not occur in this context. But in the context /V_V/ long [ŋː] is contrastive with [mː] and [nː] *(ramman, rannan, rangan* are different words). Thus, in those dialects in which [ŋː] occurs, the phoneme /ŋ/ can be postulated, albeit only on the basis of the long nasals, and e.g. *sangen* 'very' is phonemically /sɑŋŋen/. But there are also

dialects in which [ŋː] does not occur, and *sangen* is pronounced [sɑŋken]. In these dialects there is no basis for postulating the phoneme /ŋ/, and phonetic [ŋ] must be interpreted as an allophone of the phonetically nearest nasal phoneme /n/.

In those dialects in which /ŋ/ is a phoneme (in the vast majority of dialects), /ŋŋ/ sequences are often in morphophonological alternation with a /ŋk/ sequence. For example, *lanka* /lɑŋkɑ/ is the singular nominative form, *langan* /lɑŋŋɑn/ is the singular genitive form of 'thread'; the /ŋk/ ~ /ŋŋ/ alternation is an instance of qualitative grade alternation (see Chapter 1). But there are also words in which /ŋŋ/ does not participate in such alternation, e.g. *ongelma* 'problem' and *sangen*. In some recent loanwords the phoneme /ŋ/ also occurs before a heterorganic consonant, e.g. *magneetti* [mɑŋneːtːi], *kognitio* [koŋnitio], *Englanti* [eŋlɑn̪t̪i], *kongressi* [koŋresːi]. It is perhaps worth emphasising that words like *magneetti* and *kognitio*, which in e.g. English have the sequence [gn] across the first and second syllable, indeed have the sequence [ŋn] in the corresponding position in Finnish. Even speakers who have /g/ in their paradigm (see below), definitely do not pronounce *[mɑgneːtːi] or *[kognitio]. Word-internal sequences of plosive + nasal do not occur in the native vocabulary, /g/ is a marginal phoneme, and the graphemic sequence <ng> regularly represents the phonemic sequence /ŋŋ/; perhaps the reverse graphemic sequence <gn> also suggests the presence of /ŋ/ in a word to a speaker of Finnish? These are probably reasons why the example words have acquired the pronunciation they have. At any rate, the inclusion of [ŋ] suggests that the pronunciation of these loanwords is not based on that of the lending language. Also words like *Englanti, kongressi* etc. are pronounced as indicated, without a [g]. In fact, very few speakers of Finnish (L1) otherwise quite fluent in English (L2) can pronounce e.g. the two-word phrase *over England* in a native-like manner. Instead of e.g. the native RP English pronunciation [əʊvəɹɪŋglənd], assuming that the vowels are pronounced satisfactorily, three consonants are likely to reveal that the person is not an L1 speaker of RP English. Thus a Finnish L2 speaker of RP is very likely to say [əʊwə] (because Finnish does not have [v] and because [w], an allophone of /ʋ/, occurs after diphthongs similar to [əʊ] in Finnish), the speaker is very likely to omit the "linking r" between the two words (because there is no corresponding linking consonant in Finnish), and, thirdly, the speaker is very likely to say [ɪŋlənd] instead of [ɪŋglənd] (because, in Finnish, the sequence [ŋg] never occurs). In dialects in which /ŋ/ is not a phoneme, *magneetti* is pronounced [mɑŋkneːtːi], *Englanti* as [eŋklɑn̪t̪i], etc. (i.e., [ŋ] occurs as the predictable nasal before the velar plosive).

Similarly *tango* is pronounced as either [t̪ɑŋːo] /t̪ɑŋŋo/ or [t̪ɑŋko] /t̪ɑnko/, depending on the dialect.

Even in loanwords, /ŋ/ cannot occur word-initially or -finally, or as single intervocalic consonant within the word; e.g. *[ŋɑru], *[ɑʋɑiŋ] and *[kɑŋɑ] would be impossible even as loanwords.

***Group (3)**: {/d/}*

In IPA, [d] denotes a voiced alveolar plosive, but Finnish /d/ is not a plosive proper. However, IPA seems to lack a more suitable symbol, so [d] will have to do. The Finnish /d/ is apical alveolar, and the duration of its occlusion is very short, about half of that of /t̪/, ceteris paribus, see e.g. Lehtonen (1970: 71), Suomi (1980: 103). During the occlusion, the location of the apical contact with the alveoli moves forward also in vowel contexts that are in principle unfavourable to such a movement (Suomi 1998). Thus in pseudowords of the type [V_1dV_2], in which V_1 is a front vowel and V_2 is a back vowel, coarticulation would predict that the location of the contact would be retracted rather than fronted, and yet it moves forward (as it does, more expectedly, when V_1 is back and V_2 is front). That is, the fronting of the place of alveolar contact during the occlusion seems to be a special property of the Finnish /d/, which overrides the coarticulation due to vocalic context. Presumably this fronting, together with the short duration of the occlusion, contributes to /d/ being voiced: the fronting increases the volume of the cavity between the closure and the glottis, and maintains a sufficiently large transglottal pressure difference to enable voicing to continue during the brief occlusion. The fronting of the apex is reminiscent of a flap, but at the same time it is clear that the Finnish /d/ is not a flap; the fronting of the apex is not as extensive and as fast as that in a flap, and the duration of the occlusion is longer. Rather, the Finnish /d/ appears to be something half-way between a plosive (and hence obstruent) and a flap-like resonant. For these reasons, in Figure 3 below, /d/ is classified as a semiplosive.

Finnish /t̪/ and /d/ thus differ from each other in a number of respects. The places of articulation are different — /t̪/ is laminal dentialveolar and /d/ apical alveolar — and the place of /d/ is fronted during the occlusion. The duration of the occlusion of /t̪/ is roughly twice that of /d/, and finally, /t̪/ is usually voiceless, /d/ voiced. The differences in place and duration imply that the opposition is not a genuine voice opposition, as the consonants differ considerably from each other

also in other respects. Moreover, if the opposition were a genuine voice opposition, one would expect that all those speakers who have the phoneme /d/ also find it easy to pronounce [b] and [g], and to systematically distinguish between /p/ and /b/, and between /k/ and /g/. But this does not seem to be the case: speakers who have /d/ in their paradigm do not necessarily have /b/ and /g/ in their native paradigm, and they do not necessarily master the corresponding oppositions in foreign languages. On several criteria, then, /d/ is an odd one among the Finnish consonants, and hard to classify for its manner of articulation.

The synchronic phonetic and systemic oddity of /d/ has its explanation in the unusual way in which the consonant entered and spread in the language. What is now /d/ in the native vocabulary, was a few centuries ago /ð/ for all speakers. For example, the equivalent of the modern *sydän* [sydæn] 'heart' was pronounced [syðæn]. When Finnish was first written down, the mostly Swedish-speaking clerks symbolised /ð/ variably, e.g. with the grapheme sequence <dh>. When the (mostly religious) texts were read aloud, again usually by educated people whose native tongue was Swedish, <dh> was pronounced as it would be pronounced in Swedish. At the same time, /ð/ kept vanishing from the vernacular, and it was either replaced by other consonants, or simply disappeared. Today, /ð/ has vanished, and /d/ does not occur in most of the vernacular varieties in which the former /ð/ is represented by a number of other consonants or by complete loss. But /d/ does occur in modern SSF, as a result of conscious normative attempts to promote "good speaking". But even for those speakers who have /d/ in their paradigm, the consonant may not be fully stable. The speaker may use /d/ consistently in speaking in a formal register, but may replace or delete it when speaking in an informal register.

In SSF, /d/ is clearly a phoneme. In fully native words it has a rather narrow distribution, occurring as it does only word-internally in the contexts V_V (e.g. *käden* 'of hand') and /h/_V (e.g. *kahden* 'of two'). Because of this and the other restrictions, Karlsson (1983: 57) characterises /d/ a defective phoneme. In modern colloquial SSF /d/ is increasingly deleted in /hd/ sequences, e.g. *kahden* → [kɑɦen]. In the native vocabulary /d/ often alternates morphophonologically with /t̪/, e.g. *lato*: *ladon* 'barn' (nominative singular and genitive singular, respectively) but not always, as in e.g. *vihdoin* 'at last', *sydän* 'heart'. In older loanwords /d/ was always replaced by /t̪/, e.g. *tilli* < Swedish *dill*, *tuomari* < Sw. *domare*. But in newer loanwords /d/ is retained, and it thus occurs in many more positions than earlier, in very common words, e.g. *demokratia*, *indeksi*.

The phonemes /t̪/, /s/, /n/, /l/, /r/, /d/ and most of their allophones are coronal with respect to their major place feature. Let us summarise the variations in place of articulation in these consonants. In neutral contexts, e.g. in intervocalic positions, /t̪/ is laminal dentialveolar and the other coronal consonants just mentioned are alveolar, /s/ usually laminal and /n/, /l/, /r/ and /d/ apical. These can be interpreted as the respective inherent places of these consonants.

There is no major contextual variation in the place of articulation of /r/ and /d/, but /t̪/ is realised as /t/ after /s/, and /n/ and /l/ become laminal dentialveolar when next to /t̪/. The usually coronal phoneme whose allophones are not always coronal is /s/: as will be remembered from above, it often has the allophone [x] in the sequence /sr/, namely when this sequence is pronounced [xr] instead of the alternative [sɹ̥]. Otherwise, the place of articulation of /s/ does not seem to vary as a function of context.

***Group (4)**: {/f/}*

In SSF /f/ only occurs in relatively recent loanwords, such as *filmi*, *fakta*, *fasismi*, *elefantti*, *Afrikka.* In older times, a word-initial /f/ in the borrowed word was always replaced by /ʋ/ in Finnish, e.g. *vaari* 'grandfather' (< Sw. *far*), *vaara* 'danger' (< Sw. *fara*). Word-internally, /f/ was replaced by the sequence /hʋ/, e.g. *kahvi* (< Sw. *kaffe*), *sohva* (< Sw. *soffa*). Most dialects of Finnish lack /f/; in these dialects /f/ is still replaced by /ʋ/ or /hʋ/. In some other (Western) dialects that have been in contact with Swedish for a long time, /f/ does occur.

***Group (5)**: {/b/, /g/, /ʃ/}*

These consonants have entered Finnish in recent loanwords, to the extent that they can be said to occur systematically. It was mentioned earlier that /p/, /t̪/ and /k/ are occasionally voiced in rapid careless speech, and in this sense [b] and [g] do occur in all varieties of Finnish — as does [d̪], the voiced equivalent of /t̪/. But what is at issue here, is whether [b] and [g] occur systematically also in slow, careful speech, and whether they are contrastive with [p] and [k], respectively. There are many recent loanwords in which the graphemes <b> and <g> occur in the orthography, e.g. *baari*, *bakteeri*, *baletti*, *banaani*; *gaala*, *galleria*, *gamma*, *gaselli*. The question here is how such words are pronounced. In many varieties of Finnish these words begin with voiceless plosives, i.e. /p/ and /k/, respectively,

and there is thus no basis for postulating the phonemes /b/ and /g/. But in some other varieties the above words begin with voiced plosives, there are (nearly) minimal pairs like *baletti* and *paletti*, *bussi* and *pussi*, *gaala* and *kaali*, *geeli* and *keli*, and there are thus grounds for postulating /b/ and /g/ in addition to the fully native /p/ and /k/. And similarly with /ʃ/ in relation to the fully native /s/, although the number of words in which /ʃ/ potentially occurs is relatively small in comparison to those in which /b/ and /g/ may occur. Conceivably, there may be varieties in which /b/ and /g/ are phonemes, but /ʃ/ is not.

There are several factors that increase the probability of the occurrence of the phonemes /b/, /g/, /ʃ/ in the consonant inventory. Firstly, a speaker who knows foreign languages in which the corresponding phonemes occur, such as English, German, Russian or Swedish, is more likely to have these phonemes also in Finnish than a speaker who does not know such languages. Secondly, a speaker with a certain kind of social background is more likely to have /b/, /g/, /ʃ/ than a speaker with different background. High level of formal education (which usually brings about greater familiarity with foreign languages), young age, and living in urban areas all increase the probability of having these phonemes. Thirdly, speaking slowly increases the probability, because the speaker then has more time to plan the phonetic output. Fourthly, speaking in a formal register increases the probability; a speaker may have /b/, /g/, /ʃ/ when speaking in a formal register, but not necessarily when speaking in an informal one. These factors also increase the probability of /f/.

At the other extreme, then, there certainly are speakers who never have /b/, /g/, /ʃ/ in their phoneme paradigm; for such speakers e.g. *baletti* and *paletti* are homophonous, and they often use spellings such as <*proileri*>, <*krilli*> instead of the normative <*broileri*>, <*grilli*>. At the other extreme, there may be speakers who always have these phonemes; there may very well be such speakers in the large towns in Southern Finland. Between these extremes, the situation is variable and may, for a given speaker, be unstable.

It may be worth mentioning that although, due to the influence of foreign languages — nowadays overwhelmingly English — /b/, /g/ and possibly /ʃ/ are on their way to becoming regular phonemes in Finnish, there are no indications whatsoever of a similar invasion by any other non-native consonants. For example, the consonants [θ], [ð], [z] and [ʒ] constitute independent phonemes in English but do not occur in carefully spoken Finnish (but a voiced sibilant does occur as a fast-speech variant of /s/, see above) and hence, conceivably, also these

consonants could make an invasion into Finnish as emerging independent phonemes. Why this does not happen may have at least two reasons. Firstly, orthography may play a role here. Speakers of Finnish are used to the situation in which phonemic distinctions are rather systematically indicated in the orthography, and the distinctions involving /θ/, /ð/, /z/ and /ʒ/ in English are definitely much less systematically indicated in writing than those involving /b/, /g/ and possibly /ʃ/. Secondly, /θ/, /ð/, /z/ and /ʒ/ all have a very low frequency of occurrence in English, i.e. the functional load of the oppositions among these phonemes is very low. Thirdly, <z> usually represents /t̪s/ in loanwords, e.g. *zoomata* ‘to zoom’, *Fazer*.

It is our hunch that /ʃ/ may have a more marginal status than /b/ and /g/. If this hunch is correct, it too may have at least two explanations. Firstly, the spelling of /ʃ/ (in Finnish) is more variable than that of /b/ and /g/. Thus in those loanwords in which /ʃ/ is motivated, such as that for English *shock,* the orthographic forms appearing in printed texts are <sokki>, <shokki> and <šokki>. Such variability, and especially the spelling <sokki> that does not suggest any difference from the native /s/, may obscure differences between /ʃ/-words and /s/-words. Secondly, the number of words possibly containing /ʃ/ is very small.

In summary, the number of consonant phonemes in Finnish varies, depending on the variety considered, from the minimum of eleven (only the core consonants in Group 1) to the maximum of seventeen (all consonants in all Groups). That is, there is no single “correct” consonantal paradigm. Figure 3 shows the maximum system; the phonemes considered marginal are shown in parentheses. That /b/, /g/ and /ʃ/, and only these consonants, are considered marginal, reflects the authors’ estimate that a system like this might very well be the maximum system of very many speakers of the younger generation and that, even in the most formal speaking situations, there is vacillation in the phonetic realisation of these three consonants even for such speakers. This estimate does not exclude the possibility that there are speakers who have the maximum system in all speaking situations, but the estimate implies that the number of such speakers is not very large. At any rate, it seems safe to predict that the number of speakers who only have eleven consonant phonemes will diminish rapidly in the future.

	Bilabial	Labiodental	Laminal dentialveolar	Apical alveolar	Laminal alveolar	Palatoalveolar	Palatal	Velar	Glottal
Plosive	p (b)		t̪					k (g)	
Semiplosive				d					
Fricative		f			s	(ʃ)			
Glottal continuant									h
Nasal	m			n				ŋ	
Trill				r					
Lateral approximant				l					
Central approximant		ʋ					j		

Fig. 3. Classification of the Finnish consonant phonemes according to the place and manner of articulation of the respective main allophones.

4 On the phonological interpretation of the quantity opposition

Finnish is a full-fledged quantity language in that both vowel and consonant durations are contrastive, independent of each other, and independent of stress. Thus, contrastively short and long vowels can occur before and after both contrastively short and long consonants, and *vice versa,* and the contrasts exist in stressed as well as unstressed syllables. There are some restrictions to be specified below, but this is the basic principle.

According to Karlsson's (1969) "identity group interpretation" or diphonemic interpretation, contrastively long segments are interpreted as sequences of two identical phonemes, i.e. as double vowels and consonants, as against contrastively short or single ones, and diphthongs are interpreted as sequences of two different vowel phonemes. This is clearly the best interpretation. Karlsson presented a number of compelling phonotactic and morphological arguments in favour of this interpretation, and he also considered and rejected alternative interpretations suggested in the literature. One of the alternative interpretations is that long segments are considered paradigmatic phonemes in addition to the short ones, i.e. there would be a phoneme /A/ in addition to /ɑ/, a phoneme /K/ in addition to /k/, etc. (if, by convention, long phonemes are symbolised by capital letters). Recently Suomi (2008) has presented additional arguments against the paradigmatic interpretation, because in some informal discussions the diphonemic interpretation has been questioned and the paradigmatic one preferred, on very shallow criteria. One of the arguments against the paradigmatic interpretation is that, in this interpretation, all long consonants would have to be ambisyllabic. For example, the word *takka* would be phonemically /t̪ɑKɑ/, with the syllable boundary somewhere inside the /K/, since native speaker intuition definitely cannot accept the syllabifications /t̪ɑ.Kɑ/ and /t̪ɑK.ɑ/, and the word is undeniably disyllabic. If it were argued that the first part of /K/ belongs to the first syllable and the second part to the second syllable, then a question inevitably would arise as to the exact nature of these subparts of a phoneme, and of the boundary between them. If the first syllable is claimed to be /t̪ɑK/, then it would have to be said that the /K/ continues to the next syllable. Of course ambisyllabic consonants do occur e.g. in some Germanic languages (see e.g. van der Hulst, 1985), but these consonants do not have longer durations than non-ambisyllabic consonants.

Making phonetically and contrastively long consonants ambisyllabic is an otherwise unmotivated, theoretically inelegant complication.

According to strong native speaker intuition the initial syllables of *taksi* and *takka* are identical, as are the second syllables of *tuska* and *takka*. From this it follows that the first syllable of *takka* is /t̪ɑk/. The diphonemic interpretation captures this elegantly: the word is phonemically /t̪ɑk.kɑ/.

As will be explained in section 9.3 below, the duration of the second-syllable single vowel varies as a function of the weight of the initial syllable, and it is [very short] if the initial syllable is heavy. Examples of first-syllable structures after which the [very short] vowel occurs can be given phonetically as [ɑr] (as in *arki* 'workday'), [ɑrk(ː)] (as in *arkki* 'sheet of paper'), [ɑː] (as in *aamu* 'morning'), [ɑɑr] (as in *aarre* 'treasure'), [ɑɑrk(ː)] (as in *Jotaarkka*; this syllable type is rare, and the syllable in the example proper name starts with a consonant). In these notations, [k(ː)] denotes a phonetically long consonant, and the syllable affiliation of this consonant is different in the two competitive interpretations. According to the paradigmatic interpretation, these phonetic structures would be phonemically as follows (where V and C indicate long segments): VC, VCC, V, VC, VCC. Here the long consonants are problematic because, being ambisyllabic, they also belong to the next syllable. Another problem is that it seems impossible to capture the structures into a single formula; at least we cannot do it. According to the diphonemic interpretation the corresponding structures are VC, VCC, VV, VVC, VVCC. In this interpretation, the phonetically long consonants are divided into two syllables, e.g. /ɑrk.ki/ *arkki*, and there are no phonologically ambisyllabic consonants. The structures according to the diphonemic interpretation can be easily captured in a single formula: VS(C)(C) (in which S = segment). From the diphonemic structural descriptions one can also directly see how many morae they contain: as many as there are segments in a structure. In the paradigmatic structural descriptions the number of constituent morae is opaque, but VC and V are dimoraic, VCC and VC are trimoraic, and VCC is tetramoraic (unless the traditional, well motivated way of counting of morae is radically altered, which would certainly cause further problems).

These are just a couple of examples of the many structural complications that the paradigmatic interpretation would entail, complications that do not exist in the diphonematic interpretation. In brief, postulating paradigmatic phonemes for contrastively long segments would immensely complicate the description of the Finnish phonology, phonotactics and morphology. Notice that, in the

diphonematic interpretation, there is no need to talk about phonological length, or of (paradigmatic) short and long phonemes. There is only the syntagmatic difference between a single phoneme and two consecutive phonemes.

All vowel phonemes can occur as single and as double. The single – double vowel opposition is valid in any syllable, stressed or unstressed, word initially, internally and finally. Thus e.g. /pɑlɑʋɑ/, /pɑlɑʋɑɑ/, /pɑɑlɑɑʋɑ/, /pɑɑlɑɑʋɑɑ/, /pɑlɑɑʋɑ/, /pɑlɑɑʋɑɑ/ are existing, differently pronounced word forms. In some South-Western dialects, however, the vowel quantity opposition is neutralised in non-initial syllables. Hence, in the native dialect of the first author, the only quantity opposition in the above words that is realised in speech is that between the first-syllable /ɑ/ and /ɑɑ/, and thus there are, in a fully dialectal speech, only two distinct pronunciations for the six example word forms.

While all vowel phonemes participate in the quantity opposition, for consonants the situation is more complicated. Firstly, the consonants /ʋ/, /j/ and /h/, belonging to the core, only occur as single; *hihhuli* '(religious) fanatic' is the only exception. In certain dialects phonetically long [ʋː], [jː] and [hː] do occur as a result of the so-called general gemination, as e.g. in *vajaa* [ʋɑjːɑː] 'undersized' (as against [ʋɑjɑː] in SSF), but these long consonants are fully predictable as they only occur between a single vowel in the first stressed syllable and a double vowel or a diphthong in the second syllable. In these dialects, also the other consonants occur phonetically long in the same context (e.g. *sataa* [sɑt̪ːɑː] 'it rains', as against [sɑt̪ɑː] in SSF), but such phonetically long consonants are better interpreted as being lengthened single consonants rather than double consonants. The main and sufficient reason for this interpretation is the full predictability of the long duration — a phonemic distinction cannot be fully predictable — and the ensuing fact that an opposition like [ʋɑjːɑː] – [ʋɑjːɑ] is impossible (because, in the dialects under discussion, the latter form does not exist, only [ʋɑjɑ] and [ʋɑjːɑː] exist). Thus the lengthening of consonants under discussion is best analysed as a late phonetic rule that is applied in the context mentioned.

Secondly, /d/ occurs double only in recent loanwords, e.g. *addikti*, *Saddam*. But all other consonants except /ʋ/, /j/, /h/ and /d/ occur as single and as double. Thus there are minimal pairs like *rapu – rappu, kato – katto, laki – lakki, kisa – kissa, palo – pallo, kuri – kurri, tuma – tumma, vana – vanna, Kalevala – Kalevalla.* The more marginal consonants all occur as single, minimal pairs are harder to find, but here are some examples with double consonants: *rabbi* (/-bb-/), *raggari* (/-ɡɡ-/), *geišša* (/-ʃʃ-/). Of course, these minimal pairs only exist in

varieties which have the phonemes in question, and orthography is again irrelevant. But it seems, according to our informal observations, that if a variety has e.g. /b/, then it also has /bb/; unfortunately, we cannot present any supporting empirical evidence.

Thirdly, in contrast to vowels, there is no consonant quantity opposition in word initial and final positions, and only single consonants can occur in these positions (if not prohibited by further restrictions). Nor is there quantity opposition in consonant sequences, except that the true obstruents /p/, /t̪/, /k/ and /s/ can occur as single or double after nasals and the liquids /l/ and /r/, e.g. *sanka* – *sankka*, *hirsi* – *hirssi*, *pelkää* – *pelkkää*. Here as elsewhere, the double obstruents always straddle a syllable boundary.

5 Sandhi phenomena

Sandhi phenomena can be considered contextually conditioned postlexical processes that alter the shape of concatenated word forms at word boundaries and at lower-level boundaries (those between the components of compound words and morpheme boundaries), without affecting the phonemic affiliations of the segments involved. However, they can delete segments, or add aphonematic segments.

5.1 Nasal assimilation

Nasal assimilation is of course a very common sandhi phenomenon in the languages of the world. Of the Finnish nasals, only /n/ can occur word-finally in the native vocabulary. When /n/ occurs before a full word boundary or before a boundary between components in a compound word, it assumes the place of articulation of the following plosive. Thus e.g. *tytön pää* 'a girl's head' is pronounced as [t̪yt̪ømpæː], *tytön takki* 'a girl's coat' as [t̪yt̪øn̪t̪ɑkːi], *tytön kello* as [t̪yt̪øŋkelːo]. Before segments other than plosives the behaviour of final /n/ is variable. Occasionally, in very slow and formal speech, it may remain unassimilated. However, some changes usually take place. Before the resonants /j/, /ʋ/, /l/, /r/ and /m/, final /n/ may be completely assimilated to the next consonant: *järven jää* [jærʋejːæː], *järven vesi* [jærʋeʋːesi], *järven laita* [jærʋelːɑit̪ɑ], *järven ranta* [jærʋerːɑnt̪ɑ], *järven muta* [jærʋemːut̪ɑ]. Such a complete assimilation is more probable in fast and informal speech; it may also be the case that the phenomenon is more common in some dialects than in others.

Before /h/ and /s/, and possibly before /f/, final /n/ may be deleted, especially in fast and informal speech: *järven hiekka* [jærʋehiekkɑ], *järven selkä* [jærʋeselkæ], *pojan farkut* [pojɑfɑrkut̪]. However, /n/ may also be assimilated to the place of following /f/: [pojɑɱfɑrkut̪]. Before a vowel, final /n/ is retained in slow/formal speech, and in fast/informal speech its realisation is probably dialect-specific; for example, it can be deleted as an independent segment and cause nasalisation of the preceding vowel: *järven aalto* [jærʋẽɑːlt̪o], *järven yllä* [jærʋẽylːæ].

In some loanwords /m/ occurs finally, e.g. *helium*, *islam*, *slalom*. It is our strong impression that /m/ in words like these does not participate in nasal assimilation, but remains [m] in all contexts.

Thus, as a result of nasal assimilation, [m], [n̪], [n] and [ŋ] all occur word-finally. Phonemically word-final [m] may represent /n/ (as in *tytön pää*) or, more seldom, /m/ (as in *helium*), while [n̪], [n] and [ŋ] always represent /n/.

5.2 Boundary lengthening

Unlike nasal assimilation, the sandhi phenomenon discussed here is specific to Finnish. The phenomenon has been discussed under various names in Finnish linguistics, including 'final doubling', 'initial doubling', and 'final aspiration' *(loppuhenkonen* in Finnish). None of these terms is fully satisfactory, but 'boundary lengthening' may be the best compromise, and will be used here; we shall comment on these terms below.

Boundary lengthening (BL) is triggered by certain morphemes, and its effects are manifested immediately after these morphemes. BL is manifested in two different ways, depending on whether the segment following the boundary after the triggering morpheme is a vowel or a consonant. If the following segment is a vowel, then a phonetically long glottal stop appears at the boundary. For example, *Mene ulos!* 'Go out!' is pronounced as [meneʔːulos]. If the following segment is a consonant, then that consonant is lengthened, e.g. *Mene pois!* 'Go away!' is pronounced [menepːois]. In these examples the triggering morphemes are singular second person imperative forms of verbs, but there are many other types of triggering morphemes. For example, *sadekatos,* a compound consisting of *sade* 'rain' and *katos* 'shelter', thus literally 'rain shelter', is pronounced as [sɑdekːɑt̪os]. When not preceded by a morpheme that triggers BL, *katos* is pronounced with a phonetically short initial [k], e.g. in *tämä katos* 'this shelter'. Another example is the word form *rikkaillekin* 'to the rich, too', in which the enclitic *-kin* bears the meaning 'too'; this word is pronounced [rikːɑilːekːin]. Here the triggering morpheme is the allative case marker *-lle* (roughly: 'to') that precedes *-kin*. Again, when not preceded by a morpheme that triggers BL, the enclitic *-ki*n is pronounced with a short [k], as in *talokin* 'a/the house, too'. BL is thus triggered by a variety of morphemes, some of them being content words, others grammatical markers. Following Karlsson (1983), these morphemes are here referred to as "x-morphemes", e.g. *sade*^{x}. The class of x-morphemes includes certain words ending in /e/, e.g. *herne*^{x} 'bean', *koe*^{x} 'experiment', *kolme*^{x} 'three', the allative suffix *lle*^{x}*,* the adverbial derivational suffix *-sti*^{x} (*kaunis* 'beautiful', *kauniisti* 'beautifully'), the third person possessive suffix *-nsa*^{x}*,* and the singular

second person imperative forms of verbs, already mentioned, in which the morpheme signalling Singular, Second Person and Imperative Mood can be given as $Ø^{x}$ (all of these grammatical morphemes have zero expression in these particular forms). To complicate things further, there are circumstances in which an x-morpheme does not cause BL, e.g. when the allative suffix is followed by a possessive suffix: *koira+lle+ni* 'dog+to+my' = 'to my dog' is pronounced [koirɑlːeni], not *[koirɑlːenːi] (cf. *koirallekin* [koirɑlːekːin]).

Phonetically, then, if the segment following an x-morpheme is a consonant, the consonant has a duration corresponding roughly to that of a regular double (geminate) consonant within the word. And if the next segment is a vowel, a long [ʔː], whose duration also corresponds to that of a double consonant, is inserted before the vowel. Because the durations of the lengthened consonant and of the inserted [ʔː] correspond to that of a double consonant within the word, BL has been characterised as both final doubling and initial doubling, as the case may be. However, these characterisations are problematic, referring as they do to doubling (of segments). Firstly, when the segment following the x-morpheme is a vowel, e.g. in *mene ulos!* [meneʔːulos], it can be asked what in fact has been doubled here, as Finnish has no phoneme /ʔ/, and neither [ʔ] nor [ʔː] ever occur within a word (except in incompletely pronounced words, see below). Clearly, something is added here that does not belong to either of the concatenated units. Secondly, if the lengthening of the consonants following an x-morpheme is characterised as doubling (of a phoneme), then the following question inevitably arises: where should the members of the double consonant be allocated? If, for example, *mene pois!* [menepːois] is interpreted phonemically as /meneppois/, there is no satisfactory way of allocating both segments of the sequence /pp/. The interpretations /menep#pois/ and /mene#ppois/ would both immensely complicate the description of morphology and phonotactics, not to mention intuitions on how words may begin and end; for example, words would begin with double consonants only after x-morphemes. Because of problems like these, it is better to characterise the phenomenon as involving (non-phonemic) lengthening, rather than doubling. Moreover, it is preferable to state that the lengthening occurs at the boundary following x-morphemes, rather than at the beginning or end of some unit, because the prevocalic [ʔː] cannot be said to belong to either the end of the x-morpheme or the beginning of the next word. And when the initial consonant of the next word is lengthened, it is the preceding

x-morpheme that triggers the lengthening. All these things considered, boundary lengthening seems the most appropriate term.

The historical explanation of BL is that the modern x-morphemes, which all end in a vowel, once ended in a consonant that has by now disappeared. For example, the modern second person imperative form s*yö*^{x} 'eat!' was earlier [syøk], the word *kolme*^{x} 'three' was earlier [kolmet̯], *vene*^{x} 'boat' was [ʋeneh], etc. What is left of these vanished consonants is that, in the form of BL, they still emerge as consonantal material after the end of x-morphemes, when these are immediately followed by another morpheme; in utterance-final positions, BL does not surface. That BL has a historical explanation is shown e.g. by the fact that the phenomenon does not apply to recent words ending in /e/, e.g. *nukke* 'doll', *nalle* 'teddy bear'. That is, BL is no longer productive.

There is areal variation in the occurrence and scope of application of BL. In some dialects the morphemes that trigger it are more numerous than in other dialects. In the speech of the first author of this book, for example, BL is strong after singular second person imperatives of verbs, but nonexistent after words like *sade* and many other x-morphemes mentioned above, whereas for the third author BL is strong after all x-morphemes. To all appearances, BL is disappearing from the language, and very broadly speaking the disappearance proceeds from south to north.

Because it cannot be predicted on phonological grounds which morphemes trigger BL, and because the phenomenon is phonetically unstable (the lengthening may not be complete or there may be no lengthening at all), Karlsson (1983: 349) suggests that BL (which he calls initial doubling) could be characterised as a morphophonetic (rather than morphophonological) rule. We consider this an appropriate characterisation.

5.3 The glottal stop

Besides occurring between x-morphemes and vowel-initial words, [ʔ] (or a lesser degree of glottalisation) also occurs in other word-boundary positions. Itkonen (1964) reported that initial glottalisation of phonologically vowel-initial words (*alias* initial catch) is a common property of certain Eastern (Savo) dialects and some Western dialects (including many Ostrobothnian dialects). But according to Itkonen, initial catch does not occur, or is very rare, in most other dialects, including that spoken in and around Helsinki. However, Lennes, Aho, Toivola &

Wahlberg (2006), who studied informal dialogues between pairs of four female and four male young adult speakers from the Helsinki area, observed many full glottal stops in all speakers, altogether 323 tokens in four dialogues lasting 45 – 60 minutes. The speech in the Helsinki area may have changed in half a century, but there is another, more probable reason for the discrepancy. Firstly, Itkonen's results were based on dialect interviews, by himself and by others, in which the interviewer typically asks the informant short questions that usually elicit long, narrative-like replies. The speaking style in these narratives is typically very different from the very colloquial one used in the informal dialogues studied by Lennes *et al*. Secondly, Lennes *et al*. used high-quality digital recording equipment, and hence the quality of the recordings was much better than that of the old dialect recordings. Thirdly, the reported results of Lennes *et al*. are based on acoustic analysis after preliminary auditory analysis, whereas Itkonen relied on auditory analysis only. Thus it is possible that the initial glottalisation Itkonen studied occurs in narrative-like speaking styles in those dialects in which he observed it and not in other dialects, and that the initial glottalisation observed by Lennes *et al*. occurs in highly informal dialogues, in the Helsinki area and possibly in other dialects, too. That is, high informality and true dialogue (as against narration) may be factors that increase the probability of glottalisation. At any rate, the recordings and analysis methods available to Lennes *et al*. made it easier to detect glottal stops than did those available to Itkonen.

Lennes *et al*. report that glottal stops were primarily used as word-boundary signals before vowel-initial words (as ordinary initial catches or to emphasise words), during word search, and in incompletely produced words; of course, these latter occurrences are not instances of sandhi phenomena. They also mentioned that glottal stops were used for emotional emphasis. In front of vowel initial words, glottal stops tended to have roughly the same duration as other consonants in similar positions, but when a glottal stop was associated with word search or a false start, it often had a much longer duration. Glottal stops were rather common in utterance-initial position. The authors conclude that glottal stops with complete closure can be used for signalling one's intention to continue speaking, i.e. for holding the turn. We shall return to glottal stops in section 10.10.

6 Phonotactics

6.1 Vowel phonotactics

Vowel phonotactics will be discussed in two separate sections. In the first section, vowel sequences will be described from a general perspective, with a view to distinctions like monophthongs and diphthongs, number of consecutive vowels, and tautosyllabic and heterosyllabic sequences. In the second section, vowel harmony will be discussed. In both sections, ‘vowel’ is short for ‘vowel phoneme’, given the syntagmatic, diphonematic interpretation of phonetically long monophthongs and diphthongs.

6.1.1 Vowel sequences

Three kinds of sequences of two vowels within the word can be distinguished. These are double vowels, i.e. sequences of identical, tautosyllabic vowels; diphthongs, i.e. sequences of two dissimilar tautosyllabic vowels; and vowel combinations, i.e. sequences of two dissimilar, heterosyllabic vowels. Sequences of more than two vowels always contain these two-vowel sequences.

All eight vowel phonemes can occur double. The mid double vowels /ee/, /øø/ and /oo/ are lexically less frequent than the others. This is because of a sound change that took place in Early North Finnic, in which */ee/, */øø/ and */oo/ became the diphthongs /ie/, /yø/, /uo/, respectively. For example, **mees* ‘man’ became *mies,* **töö* ‘work’ became *työ,* and **tooli* ‘chair’ became *tuoli;* in Estonian, which is closely related to Finnish, the corresponding changes did not take place. Later, however, new /ee/, /øø/ and /oo/ sequences have entered the language through borrowing on the one hand, and through sound changes in the native vocabulary, on the other. For example, *teeri* ‘black grouse’, *insinööri* ‘engineer’ and *moottori* ‘motor’ have been borrowed after the sound change was productive. In many native words, consonants have been deleted between former single vowels, e.g. former *kätehen* ‘to a/the hand’ has become *käteen.* Despite these later developments, then, mid double vowels are still less frequent than the other double vowels.

According to the traditional classification, there are 18 diphthongs. Of these, fifteen end in a high vowel quality: /ei/, /yi/, /øi/, /æi/, /ɑi/, /oi/, /ui/; /iu/, /eu/, /ɑu/, /ou/; /ey/, /iy/, /øy/, /æy/; and three end in a mid vowel quality: /ie/, /yø/,

/uo/ (these latter three are the diphthongs that were monophthongs in Early North Finnic). The diphthongs /yi/, /øi/, /iy/ and /ey/ are rare in word-initial syllables and in free morphemes generally, examples are *lyijy* 'lead', *söi* '(he/she) ate', *leyhytellä* 'to fan' and *Kiysaari* (proper name). The rarity of these diphthongs can be formally explained by stating that they violate a restriction that may be called labial harmony. This harmony only applies to the vowels /i/, /y/, /e/ and /ø/, and the restriction is that a diphthong consisting of these vowels has to be either rounded throughout or unrounded throughout. The diphthongs that do not violate labial harmony, e.g. /ei/ (unrounded throughout) and /øy/ (rounded throughout), are frequent.

The diphthongs discussed above occur in SSF and in many dialects. In some Eastern Finnish dialects SSF diphthongs correspond to monophthongs, and vice versa. E.g. SSF *maa* 'land' is *mua* in some Savo dialects, and *kauhea* 'terrible' is *kaahee.* In many dialects the diphthongisation process of the mid double vowels that started in Early North Finnic has gone further, so that /ie/, /yø/, /uo/ have become /iɑ/, /yæ/, /uɑ/, respectively; in fact, because of vowel harmony, the back harmonic /iɑ/ has the front harmonic variant /iæ/.

According to the traditional classification, there are 20 vowel combinations, all with a syllable boundary between the two vowels: /i.ø/, /i.æ/, /i.ɑ/, /i.o/; /e.ø/, /e.æ/, /e.o/, /e.ɑ/; /y.e/, /y.æ/; /ø.e, /ø.æ/; /æ.e/, /æ.ø/; /ɑ.e/, /ɑ.o/; /o.e/, /o.ɑ/; /u.e/, /u.ɑ/. Three of these, however, must be considered marginal, namely /y.æ/, /ø.æ/ and /u.ɑ/. In all vowel combinations, the second member is never high; it is either mid or low. Of the 17 non-marginal combinations 14 occur across a word's first and second syllable, while all 17 occur later in the word.

Implicit in the classification of sequences of two dissimilar vowels into diphthongs on the one hand and into vowel combinations on the other, is the claim that a given sequence can belong to only one of the two classes. The situation is not so straightforward, however. For example, speakers disagree among themselves as to whether words like *pian* 'soon', *tae* 'guarantee' or *teos* 'work' are mono- or disyllabic, or whether *oikeus* 'justice' or *talous* 'economy' are di- or trisyllabic. In each case, the uncertainty concerns the sequence of two vowels. Thus e.g. /eu/ and /ou/ can be judged to be heterosyllabic sequences in some words by some speakers, while they are unquestionably tautosyllabic for all speakers in words like *leuka* 'chin' and *koulu* 'school'. Thus while there are no ambisyllabic consonants in Finnish, there is ambivalence concerning the syllabic

division in some vowel sequences; Häkkinen (1978) has studied this problem experimentally.

There are also sequences of three and four vowels, decomposable into combinations of shorter sequences. For example, *ai.emmin* 'earlier', *hau.is* 'biceps', *vaa.oissa* 'in a pair of scales', *tai.oit* 'you conjured'. Altogether, then, Finnish allows very long vowel sequences within the word. There can be very many structurally different vocalic portions, such as single vowels, double vowels, diphthongs, vowel combinations, and combinations of these (up to four vowels in a row). This, together with vowel harmony, sets Finnish apart from many other languages.

6.1.2 Vowel harmony

With respect to vowel harmony, the Finnish vowels belong to one of three classes: the front harmonic /y/, /ø/, /æ/; the back harmonic /u/, /o/, /ɑ/; and the harmonically neutral /i/, /e/. Notice that the front and back harmonic vowels correspond pairwise to each other with respect to vowel height and rounding: /y/ to /u/, /ø to /o/, and /æ/ to /ɑ/, and that the harmonically neutral vowels are front peripheral vowels. The major restriction is that, within an uncompounded word, vowels from the front harmonic and back harmonic classes cannot co-occur, while harmonically neutral vowels can co-occur with vowels from both harmonic classes. Thus there are words like *kylä* 'village' (only front harmonic vowels), *talo* 'house' (only back harmonic vowels), and *isä* 'father', *kirja* 'book', *kesä* 'summer', *kello* 'clock' (a mixture of harmonic and neutral vowels). Vowel harmony does not apply across the boundary between the components of compound words: *iso+isä* 'grandfather', *kesä+loma* 'summer vacation'.

All suffixes that contain harmonically non-neutral vowels have both a front harmonic and a back harmonic variant. For example, the singular inessive of *talo* is *talo<u>ssa</u>* 'in a/the house', that of *kylä* is *kylä<u>ssä</u>* 'in a/the village'. The stem of the word determines the harmony class of the suffixes: if the stem is back harmonic, also the suffixes are back harmonic, otherwise the suffixes are front harmonic. Thus the suffixes are front harmonic if the stem is front harmonic (as in *kylässä),* but also if the stem is harmonically neutral *(peli* 'play', *pelissä* 'in a/the play').

Words of the type *isä*, *kesä*, *heinä*, *etelä* (neutral + front harmonic vowels) are older than words of the type *kisa*, *kela*, *velka*, *hiekka* (neutral + back harmonic vowels). In older times, after initial syllables containing only (combinations of) /i/

and/or /e/ but no other vowels, only front harmonic but not back harmonic vowels occurred in later syllables (as in *isä*). Later, words of the type *kisa* started to gain ground, and nowadays the old harmonic pattern (neutral + front harmonic vowels) has lost its productivity. In new words, usually only back harmonic vowels occur after neutral vowels in the first syllable, i.e. back harmonic vowels are now productive in this position. This can be seen e.g. in colloquial short forms of longer words, such as *eka, Hesa* and *vimppa,* from *ensimmäinen* ‘the first one’, *Helsinki* and *viimeinen* ‘the last one’, respectively. The longer words are front harmonic, e.g. the singular inessive forms are *ensimmäisessä, Helsingissä* and *viimeisess*ä (i.e. the ending is *-ssä,* not *-ssa).* And yet the short forms are back harmonic: *ekassa*, *Hesassa*, *vimpassa*. It seems that new words like **ekä*, **Hesä* and **vimppä* no longer arise, and in fact they sound ungrammatical to a native speaker. In spite of this, the old words following this pattern show no signs of change.

There are three common, fully native words that are exceptions to vowel harmony. The singular partitive forms of *meri* ‘sea’ and *veri* ‘blood’ are *merta* and *verta,* respectively, i.e. they are back harmonic even though all other singular and plural inflected forms in the inflectional paradigm are front harmonic, e.g. *meressä* ‘in sea’ and *veressä* ‘in blood’. The third exception is *tällainen* ‘like this’ that contains both front and back harmonic vowels. The explanation is historical: *tällainen* is a merger of the words *tämän* ‘of this’ and *lainen* ‘like’.

There are a large number of recent loanwords that violate vowel harmony, e.g. *dynamiitti, marttyyri, hypoteesi, symboli, Hyla*. In former times loanwords violating vowel harmony were always adapted to conform to it, e.g. *tyyny* ‘pillow’ < Sw. *dyna, myssy* ‘cap’ < Sw. *mössa, ryöväri* ‘robber’ < Sw. *rövare,* but this no longer seems to take place, although some individual speakers may still apply the adaptation. The pronunciation of the words violating vowel harmony usually causes no problems, and thus there seems to be no hard pressure today towards adapting them to the old pattern. Some individual words cause problems, however; *olympialaiset* ‘the Olympic games’ is a notorious example, it is very often pronounced [olumpiɑlɑiset̪], a pronunciation that is almost as often reproached; one gets the impression that sports fanatics consider this pronunciation an act of sacrilege. In words violating vowel harmony, there is vacillation in the choice of the suffixes containing harmonic vowels: for example, one may observe either the form *parametrissa* or the form *parametrissä,* in both speech and writing.

In many languages, the number of vowel phonemes that can occur in the stressed syllable is larger than that in unstressed syllables. In certain varieties of Swedish, for example, nine short and nine long vowels can occur in the primarily stressed syllable, whereas in other syllables only the short vowels can occur. In syllables preceding the primarily stressed syllable, seven vowels can occur, and in syllables following the primarily stressed one, the number of possible vowels diminishes as distance from the stressed syllable increases. Thus in the syllable following the primarily stressed one, seven vowels are possible, in the next syllable five, and in later syllables only two (Garlén 1988). This means that the listener need not distinguish between as many vowel qualities in secondarily stressed and unstressed syllables as in primarily stressed ones. Finnish has no restrictions on the occurrence of vowels dependent on stress in the sense that all vowels, both single and double, can occur in any syllable of the word. But it follows from vowel harmony that in Finnish, too, the selection of vowels is in a way smaller in syllables following the first, primarily stressed syllable than in the primarily stressed syllable itself. For if a word's first vowel is e.g. /ɑ/, then it is highly improbable that /y/, /ø/ or /æ/ should occur later in the word; this can be the case only in relatively recent loanwords like *kanyyli, manööveri* and *afääri.* Thus, excluding recent loanwords, it is the case that a front harmonic first vowel of a word can be followed by only front harmonic and harmonically neutral vowels, and correspondingly a back harmonic first vowel of a word can be followed by only back harmonic and harmonically neutral vowels.

If a word's first vowel is a harmonically neutral vowel, then too only five different vowels can occur later in the word, again excluding recent loanwords. In this situation the listener just cannot yet know, when hearing the first vowel, whether the harmonic vowels possibly occurring later in the word are front harmonic or back harmonic. If a word's first vowel is /i/ or /e/, then /i/ and /e/ can occur later in the word, as can vowels either from the class /y ø æ/ or from the class /u o ɑ/, but not from both of the latter classes. Thus there are words like *ihminen* 'human being', *ikävyys* 'dullness' and *ihana* 'lovely', but **ikavyys, *ikävuus* as well as **ihanä, *ihäna* would be impossible in the native vocabulary. In other words, assuming that the listener correctly recognises a word's first non-neutral vowel, only five vowels have to be distinguished in later syllables, not eight as in the primarily-stressed syllable. Since, in older times, loanwords violating vowel harmony were fully adapted to Finnish sound structure, what has just been said held good without exception.

Suomi (1983) proposed that the original motivation for vowel harmony is a perceptual one, to facilitate word recognition by decreasing the number of vowel qualities that have to be distinguished in syllables following the initial syllable; this idea has later been adopted in explaining e.g. rounding harmony (Kaun 2004). Thus Finnish vowel harmony results in the same state of affairs as the vowel occurrence restrictions related to stress result in many other languages: the number of possible vowel contrasts is smaller in non-primarily stressed syllables than in primarily-stressed ones.

Even today, if there is a harmonic mismatch between consecutive vowels separated by any number of consonants, i.e. if the in this sense consecutive vowels belong to opposing harmony classes (the former to front harmonic and the latter to back harmonic, or *vice versa),* it is highly probable that there is a word boundary between the vowels. For example, in the sequence /-uCy-/ it is very probable that /u/ and /y/ do not belong to the same word. But vowel harmony can make no prediction concerning the presence or absence of a word boundary in e.g. the sequences /-uCu-/ and /-yCy-/: for all we know, the vowels may or may not belong to the same word. Suomi, McQueen & Cutler (1997) demonstrated that Finnish listeners can exploit harmonic mismatch information in an on-line speech segmentation task. For example, listeners found it easier to detect words like *hymy* at the end of the nonsense string *puhymy* (where there is a harmony mismatch between the first two syllables) than in the string *pyhymy* (where there is no mismatch). Similarly, *palo* was detected easier in *kypalo* than in *kupalo*.

Hakulinen (1961) insists that "a dominant principle in the formation of words out of the sounds of Finnish is the *avoidance of all phonemes which are difficult to articulate or which require a comparatively tensed use of speech organs*" (p. 6; emphasis in the original), that "the most notable application of this principle is the phenomenon of vowel harmony" (p. 7), and that "admittedly the pronunciation of a phoneme contrary to the rule of vowel harmony does not cause any difficulty to those who speak several European languages — even so closely related language as Estonian has words like *häda* (F. *hätä,* trouble) — but it is nonetheless a physiological fact that the observance of vowel harmony constitutes the same type of avoidance of articulatory effort as the phenomenon of assimilation" (p. 7). We leave it to the reader to judge the relative merits of the two, mutually irreconcilable explanations of the motivating causes of vowel harmony (but we point out that e.g. the Finnish trill [r] is notoriously difficult for many foreigners, as is the quantity opposition).

6.2 Consonant phonotactics

6.2.1 Word-initial consonant sequences

It is here that misconceptions of Finnish sound structure are perhaps greatest in the linguistic literature. Hakulinen (1961) writes that “a second application of the principle of avoidance of difficulty in articulation is observable in the peculiarity — also probably inherited from Finno-Ugric — that a *syllable* (and consequently a word) *never begins with a cluster of consonants”* (p. 7, emphasis in the original). This strict and unconditional statement was definitely an anachronism already half a century ago. As will be shown presently, word-initial consonant sequences do occur. To make things completely clear, when we say below that such and such consonant sequence occurs under such and such conditions, we of course mean that the sequence of consonants is actually pronounced in speech; that is, there are borrowed words that are normatively written in a certain way, but their actual pronunciation may vary, and we only refer to the pronunciation unless explicitly stated to the contrary.

Let us first look at word-initial single consonants, or C “sequences”. In older times, what Hakulinen writes above was true, and also all loanwords were adapted to this pattern, e.g. *ranta* ‘shore’ < Sw. *strand, peli* ‘play’ < Sw. *spel, ruuti* ‘gunpowder’ Sw. < *krut.* But today the situation is very much different. Of the fully native core consonants (Group 1 above) all occur word-initially, and /d/ occurs in this position in very common loanwords, e.g. *demokraatti, desimaali, devalvaatio, diktaattori.* The words are pronounced with an initial /t̪/ by some older speakers of some varieties, but it is very difficult to believe that there would be speakers, without speech disorders, in their twenties who do not have /d/ in this position when they speak SSF. Also /f/, /b/, /g/ and /ʃ/ occur word initially in those varieties and registers that have them, e.g. *fonetiikka*, *baari*, *geeni*, *šokki.* These words too are not pronounceable with these consonants by all speakers.

Word-initial CC sequences can be divided into three structural groups. The first group consists of five plosive + liquid sequences: /pl/ (*planeetta*, *plussa*), /pr/ (*prosentti*, *presidentti*), /t̪r/ (*traktori*, *tropiikki*), /kl/ (*klinikka*, *klubi*) and /kr/ (*kriisi*, *kruunu*). These sequences are also common in native descriptive words, e.g. *prätkä* ‘motorbike’, *plörö* ‘liquor’ (both words have humorous connotations). Of the theoretically possible word-initial plosive + liquid sequences */t̪l/ is clearly prohibited; this is very probably due to the fact that this sequence is

prohibited word-initially also in the languages from which words have mostly been borrowed to the language (formerly Swedish, now English). The second group of word-initial CC sequences consist of /s/ + /p/, /t̪/ or /k/, e.g. *spanieli, spekulaatio, sponsori; statisti, stereo, stipendi; skaala, sketsi, skootteri.* The third group is structurally a mixed one, and the degree of domestication of the sequences is variable. Examples are *psykologi*, *tsaari*, *snobi*, *draama*, *fraasi*, *britti*, *gramma* and *flunssa* 'flu'. All three groups of CC sequences may be simplified in some varieties; when this happens, it is the second consonant that remains (e.g. s*panieli* > *panieli),* and if there is no simplification, the non-core consonants may be replaced by their fully native phonetic neighbours. Thus e.g. SSF *gramma* may be *ramma* or *kramma* in some varieties.

There are also word-initial CCC sequences, but they are rare in the vocabulary. Among the most common ones are /spr/ and /st̪r/ — especially the latter occurring in very frequently used words —, e.g. *sprii*, *sprintteri*, *strategia*, *stressi*. Also these sequences are simplified in many varieties, e.g. *stressi* > *ressi*; to what extent theoretically intermediate forms such as *tressi* occur, is unclear to us. But what is clear is that the last consonant in consonant sequences is always retained: e.g. *stressi* does not become *sessi* or *tessi*.

In the past, then, Finnish tolerated only singleton consonants at word onset, and all borrowed words were adapted to this pattern. Today, the situation is very much different, and sweeping generalisations *à la* Hakulinen (1961) reflecting the old pattern are simply untenable. At present, there is fluctuation in the way longer word-initial consonant sequences are pronounced. An intelligent guess might be that, in the future, these sequences will be fully established in the language. Together with the inventory of consonant phonemes, the phonotactics of word-initial consonants is an area in which foreign influence is very conspicuous.

6.2.2 Word-internal consonant sequences

As C sequences, all consonants except /ŋ/ can occur (to the extent that a given consonant occurs in a variety). CC sequences can be divided into double consonants or geminates and sequences of two dissimilar consonants; in all word-internal CC sequences, there is a syllable boundary between the two consonants. The double consonants /dd/, /bb/, /ɡɡ/, /ff/ and /ʃʃ/ only occur in recent loanwords (examples were given above), and /hh/ (with the exception of *hihhuli* mentioned earlier), /jj/ and /ʋʋ/ do not occur at all. In CC sequences consisting of

dissimilar consonants, in the native vocabulary /d/ only occurs in /hd/ and /ŋ/ only in /ŋk/, but in recent loanwords both consonants also occur in other contexts, as in e.g. *kandi(daatti), kognitio* [koŋnit̪io]. Otherwise, the most important restrictions on word-internal CC sequences are the following:

1. A nasal cannot follow a plosive. There are exceptions in recent loanwords, e.g. *hypnoosi, luutnantti, rykmentti, tekniikka.*
2. /l/ and /s/ cannot be followed by /r/. Presumably both restrictions have a similar articulatory motivation: /r/ requires an activity of the tongue blade that is difficult to accomplish after /l/ and /s/. Exceptions: loanwords like *Kilroy*, *Israel*. The reverse sequences /rl/ and /rs/ in turn are common. Recall from above that the sequence /sr/ (which is common across a word boundary and across the boundary between the components of compound words) has the alternative pronunciations [xr] and [sɹ̥]. Thus in the /sr/ sequence both phonemes are realised, but in such a way that a sibilant is not followed by a trill. In fact, this is a kind of phonetic sandhi restriction: the phonemic affiliations of the consonants do not change, but if the occurring allophone of one of the consonants is a sibilant or a trill, then the other consonant must be realised by an allophone that does not occur elsewhere.
3. A nasal cannot be followed by a liquid. Exceptions: *Venla*, and loanwords, e.g. *vänrikki, genre, Englanti* [eŋlɑnt̪i].
4. A central approximant cannot be followed by a consonant, as central approximants cannot occur in the coda position. Exceptions: the loans *sovhoosi* and *klovni*.
5. An obstruent cannot be followed by /h/, except across a morpheme boundary. Monomorphemic native words of the type **lathi* do not occur, but there are word forms like *saat+han* ‘you do get, don’t you’, in which *-han* is a pragmatic enclitic; the translation is only approximate. There are loanwords spelled with a <th> sequence, such as *menthol(i), python,* but these words are at least very often pronounced without /h/.
6. A plosive other than /t̪/ cannot be followed by a central approximant. Thus there are native words like *latva* and *patja*, but none like **lakva*, **lapva*, **lakja* etc. In loanwords, however, at least /kʋ/ occurs, as e.g. in *akvaario*, *ekvivalentti*.
7. A nasal cannot be followed by a heterorganic consonant. There are true exceptions in loanwords well domesticated into Finnish, e.g. *linja* ‘line’,

limsa (colloquial for) 'lemonade', *magneetti* [mɑŋneet̪ːi], as well as in fully native-sounding proper names such as *Anja, Sonja, Jämsä, Komsi.* The sequence /nh/ is very common, but /h/ in this context is realised as the glottal allophone [h], and [h] has no supraglottal place of articulation. Therefore, if 'heterorganic' is interpreted to refer to a difference in supraglottal place of articulation, then /n/ and /h/ are not heterorganic in the sequence /nh/, and then this sequence is not an exception to the generalisation.

8. A labial plosive cannot be followed by a non-labial plosive, and a non-labial plosive cannot be followed by a labial plosive. Thus, in the native vocabulary, the sequences /pt̪/, /pk/, /t̪p/ and /kp/ do not occur. There are exceptions in recent loanwords at least as concerns /pt̪/, e.g. in *apteekki*, *kapteeni*, *optimistinen*. Across a morpheme boundary, /t̪p/ also occurs in fully native words, e.g. in *olet+pa* 'thou certainly art'; *-pa* is another pragmatic enclitic, and the translation given is only approximate.
9. A velar plosive cannot be followed by a dentialveolar plosive. In loanwords, however, the sequence /kt̪/ is quite usual: *traktori*, *aktiivinen*, *taktiikka* etc.

It follows from restrictions (8) and (9) that the only sequence of two plosives occurring in fully native monomorphemic words is the very common /t̪k/.

The restrictions mentioned above do not cover all non-occurring but theoretically possible word-internal CC sequences, and further restrictions could be formulated. It is not always clear whether a given sequence should be considered permitted or prohibited. For example, the female name *Venla* is the only uncompounded word in the language in which the sequence /nl/ occurs, yet the sequence is fully pronounceable and sounds fully native (the Germanic *Vendela* is the probable source of this name). Of course *Finland* is another proper name in which this sequence occurs, but this word is much less domesticated because of its initial and final consonants.

There are many word-internal CCC sequences; in such sequences, there is always a syllable boundary before the last consonant. The largest number of CCC sequences is found across the boundary between the first and second syllable. In other than the most recent loanwords the first consonant is a liquid or a nasal, the other two are true obstruents (/p/, /t̪/, /k/ or /s/), e.g. *helppo*, *tarkka*, *kurssi*, *palsta* and *tontti*. CCC sequences ending in a single /s/ often convey special, informal connotations, e.g. *lonksua* 'to rattle', *rempseä* 'easy-going', *kampsut* 'belongings' (the translations do not convey the connotations). Among the fully native words,

horsma ‘willow’ is exceptional in having a non-obstruent as the last consonant of the CCC sequence. In some colloquial shortened words with informal connotations the three-obstruent sequence /t̪sk/ occurs: *jätski* ‘ice cream’ (from *jäätelö), matsku* (from *materiaali)*, *motskari* ‘motorbike’ (from *moottoripyörä*).

In recent loanwords a large variety of CCC sequences occur, which reflect the corresponding sequences in the lending languages: *teksti*, *spektri*, *impressio*, *röntgen*, *mordva.* CCCC sequences are less frequent and only occur in recent loanwords, e.g. *instrumentti*, *abstrakti*, *ekspressiivinen*, *hamstrata* ‘to hoard’. In these CCC(C) sequences, too, marginal consonants are replaced by their fully native phonetic neighbours in many varieties. To our knowledge, word-internal CCC(C) sequences in loanwords are not regularly shortened in any variety, nor is there any tendency to insert epenthetic vowels to simplify the sequences, as happens in some other languages.

6.2.3 Word-final consonant sequences

In fully native words, only the consonants /t̪/, /s/, /n/, /l/ and /r/ can occur word-finally, e.g. *olut* ‘beer’, *vieras* ‘guest’, *nainen* ‘woman’, *manner* ‘continent’, *sävel* ‘tune’. Of these, however, /l/ and /r/ are very rare word-finally. If full, non-reduced word forms are considered, then Finnish has practically no word-final CC sequences (or longer ones). There are a couple of onomatopoetic interjections like *poks, rits, plumps,* and a couple of loanwords: *morjens* ‘hello’ (informal) and *preesens* ‘the present tense’. But in many dialects many word-final vowels (and some other segments) are regularly deleted (in comparison to SSF), and this also happens in colloquial, informal versions of SSF, and in such varieties word-final CC sequences are very frequent, e.g. (the vowels in parentheses are deleted in these varieties): *miks(i)* ‘why’, *yks(i)* ‘one’, *kenelt(ä)* ‘from whom’, *meneks* ‘are you going’ (from the full form *menetkö sinä)*.

Borrowed words that end in one or more consonants in the lending language are usually adapted to Finnish phonotactics by adding a vowel to the end. This has happened in the past, and it is happening today. Examples of old loans are *masto* ‘mast’ < Sw. *mast, syltty* ‘brawn’ < Sw. *sylt, santa* ‘sand’ < Sw. *sand*. In the past, any vowel could be added to the end of the borrowed word (observing vowel harmony, however), but now the added vowel is invariably /i/, as in *kurssi*, *presidentti* and *trendi* (but there are at least two recent slang word exceptions to this generalisation, namely *stara* ‘(pop) star’ and *handu* ‘hand’). In this way, the

originally word-final consonant (sequence) is made word-internal. Usually /i/ is also added to words that would otherwise end in /t̪/, /s/, /n/, /l/ or /r/ and which, as such, would be consistent with Finnish phonotactics, e.g. *analyysi, mikrofoni, konsuli*, *printteri*. But there are also some established loanwords to which /i/ has not been added, e.g. *anis*, *tennis*, *karies*, *neon*. Finally, there are words that do not have final /i/ in SSF, but do have it in some other variety, e.g. *nailon(i)*, *Eeden(i)*, *röntgen(i).* In loanwords that are monosyllabic in the lending language, a tendency towards disyllabicity also promotes the addition of /i/ to the end, e.g. *pop* > *poppi*, *deck* > *dekki*, and this also happens in words that would otherwise end in /t̪/, /s/, /n/, /l/ or /r/: *bit* > *bitti*, *mail* > *meili*, *gel* > *geeli* etc.

Usually, if a word is monosyllabic in the lending language, if it ends in a singleton consonant, and if the vocalic portion is represented by a single vowel in Finnish, the word-final consonant is doubled, e.g. *poppi, dekki* (already mentioned), *bussi, pinni, rommi.* This is not always the case, however. Thus English *fan* 'enthusiastic follower' has been established as *fani* (not *fanni)*. It also seems that if the final consonant of the borrowed word is a voiced stop in the lending language, it is not doubled. Thus *loki* 'record of a ship's daily progress' is a relatively old loan from English in which /k/ has replaced /g/, English *grog* has been later borrowed as *grogi* (pronounced /grogi/ by many), Swedish *glögg* (another sort of alcoholic drink) as *glögi*, and the recent loan *blog* 'net diary' seems to be pronounced /blogi/ (not /bloggi/).

All example words in the preceding two paragraphs are uninflected nouns (i.e. singular nominative forms). When inflected, then suffixes are added to them, and thus consonants that are word-final in singular nominative, are not word-final in the inflected forms. For example, *tennis* is *tennistä* in the partitive, *tenniksen* in genitive, etc. As /t̪/ and /n/ are the only consonants that can occur suffix-finally, their proportion of word-final consonants must be very high; as already mentioned above, /l/ and /r/ are very rare word-finally.

Finnish thus actively avoids, at the end of unreduced forms of words, sequences of two or more consonants, and to a considerable extent also singleton consonants. The means to accomplish this are somewhat variable, as was shown for words ending in /t̪/, /s/, /n/, /l/ or /r/. There is a conspicuous difference between what is happening, as a result of extensive borrowing, to word-initial and word-internal consonant sequences and to word-final ones. At word onset and word-internally, increasingly complex sequences are clearly gaining ground, while at word offset no corresponding change is visible because the language

makes use of ways to avoid word-final consonant sequences. Languages like Spanish avoid word-initial consonant sequences by inserting an epenthetic vowel to word onset: Finnish does not do this, but instead inserts an epenthetic vowel to word offset to avoid word-final consonant sequences.

6.3 Restrictions on #CV, #VV and #(C)VVCC sequences

The few restrictions on the combinability of consonants and vowels (in this order) are discussed in this section; all of these restrictions can be stated with reference to word onset. The #CV restrictions concern three word-initial CV sequences, and the #VV restrictions three word-initial VV sequences. These restrictions concern central approximants as well as high and mid vowels. One of the three #CV restrictions is that the sequence /ji/ is prohibited word-initially. All words mentioned in the dictionary Nykysuomen Sanakirja ('Modern Finnish Dictionary', hereafter NS) beginning with /ji/ are loans: *jiddiš, jigi, jiikata, jiiktouvi, jiina, jiirata, jiiri, jiki, jiujitsu.* These words are all very infrequent, and many are unknown to most speakers of Finnish. Another, weaker restriction prohibits the sequence /je/ word-initially, the words mentioned by NS are *jee, jeep, jeeveli, jefreitteri, jehu, jekku, jen, jenka, jenkka, jenkki, jennykone, jeremiadi, jermu, jermuilla, jes, jestas, jesuiitta, jetoni* ja *jetsulleen,* of which the majority are recent loanwords, and the majority of the native ones have special connotations. It is clearly a question of restrictions on word-initial sequences, as elsewhere both /ji/ and /je/ are quite frequent in native words, e.g. *laji* 'species', *koje* 'devise'.

Finnish also has conspicuously few native words beginning with /ii/ and /ie/. For /ii/ NS mentions the five nouns *iikka, iileskotti ~ iiliskotti, iili(mato) ~ iiliäinen, iippa,* all of which are rare and four of which convey special connotations, and there are a handful of proper names like *Ii, Iiro, Iittala, Iivonen.* Words beginning with other double vowels are much more numerous; words beginning with the mid double vowels are also relatively rare in this and other positions but this circumstance has an independent reason, the historical change of */ee/, */øø/ and */oo/ to /ie/, /yø/, /uo/, respectively, as discussed above. For /ie/ NS mentions only two nouns, and there do not seem to be any proper names. In contrast, words beginning as /Cii/ and /Cie/ are a legion, as long as the C is not /j/. So there seems to be a restriction on both *#/ii/ and *#/ie/. Acoustically and perceptually, the word initial sequences /ji/ and /ii/ are very much like each other, and so are word initial /je/ and /ie/. Consequently, then, the #CV restrictions and

the #VV restrictions clearly have a common functional motivation: to avoid word initial sequences that are easily confusable.

The third #CV restriction prohibits the word-initial sequence /ʋu/, except when the extension of the sequence is /ʋuo/, the sequence that before the change in Early North Finnic discussed above was */ʋoo/. It can thus be stated that there is a restriction on *#/ʋu/, with the exception of the very common #/ʋuo/. The words violating this restriction in NS are *vualee* (or *voile), vuitti, vulfeniitti, vulgääri, vulkaani* (and derivatives), *vulmahti, vulpiinihappo, vulsti, vulva, vunteerata, vunukka.* These are either rare dialectal words *(vuitti, vunukka)* or hardly known loanwords.

Also words beginning with /uo/ seem to be rare. NS mentions only one commonly known noun *(uoma)* and two dialectal words, and there are a handful of proper names like *Uolevi, Uosukainen, Uoti.* Thus the third #VV restriction is *#/uo/. Also the restrictions *#/ʋu/ and *#/uo/ have an obvious common functional motivation: to avoid word initial sequences that are easily confusable; /ʋu/ would be confusable with /uu/, /uo/ with /ʋo/.

There are asymmetries in the restrictions *#/ji/ and *#/je/, *#/ii/ and *#/ie/, *#/ʋu/ and *#/uo/. Firstly, there is an asymmetry in that the restrictions involving front vowels are more numerous (*#/ji/ and *#/je/, *#/ii/ and *#/ie/) than those involving back vowels (*#/ʋu/ and *#/uo/). Secondly, there is an asymmetry in that while #/ii/ is prohibited, #/uu/ is not. These asymmetries are most probably due to the fact that /j/ and /i/ are more alike phonetically than are /ʋ/ and /u/: the greater the risk of confusion, the stricter the restrictions.

There are also #(C)VVCC restrictions. There are obviously no restrictions on VC sequences, i.e. restrictions on sequences of a single vowel and a single consonant anywhere in the word, while there are restrictions on #(C)VVCC sequences: that is, all theoretically possible combinations of VV and CC are not possible in the first syllable. It does not seem to matter which vowels make up the VV part of the #(C)VVCC sequence — apart from vowel harmony —, but not all consonants can make up the final CC part. These restrictions are in addition to those applying to all word-internal CC sequences discussed above, i.e. there are further restrictions on CC after #(C)VV that are not found in other word positions. The most important of the CC restrictions in the context #(C)VV__ are the following; the restrictions apply to words irrespective of their morphological structure, unless otherwise stated:

1. A resonant cannot be followed by a heterorganic consonant, e.g. **tuulka, *saarpo.* The sequence /rm/ is a common exception, e.g. *kuorma* ‘load’, *käärme* ‘snake’.
2. The glottal /h/ cannot be followed by a resonant, e.g. **taahna,* **suohli*. This restriction is not valid in some dialects in which e.g. *haahmo* (for SSF *hahmo)* ‘figure’ occurs.
3. A non-coronal resonant cannot be followed by another consonant, e.g. **soimpo* and **laanki* [lɑːŋki]. However, if a morpheme boundary intervenes the two consonants, the restriction does not apply, e.g. *saa+n+ko* [sɑːŋko] ‘do I get?’ (= get + I + question). A coronal resonant followed by another consonant in turn is a common sequence after #(C)VV, e.g. *vienti, saarto.*

For more details of the #(C)VVCC restrictions see Suomi (1990).

7 Syllable and mora structure

There are ten types of syllables that occur in fully native words. They can be characterised as the basic syllable types, and their structure can be described by the template (C)V(S)(C) in which "S" refers to a segment, either V or C, and in which each segment is a phoneme, given the syntagmatic interpretation of quantity. A minimal syllable thus consists of a single vowel; the syllable nucleus is always a vowel. In Finnish, the syllable nucleus is the syllable's first mora, and every phoneme segment following in the same syllable constitutes an additional mora. The ten basic syllable types are given below in Table 1, with information of the proportion of each type of all basic types according to Häkkinen (1978), and structural information.

Table 1. The basic syllable types, their frequency of occurrence, example words, weight, number of morae and the structure of the rhyme.

Syllable type	Proportion	Example	Weight	N of morae	Rhyme
CV	40.4	***ta***.*lo*	light	1	V
CVC	27.5	***tas***.*ku*	heavy	2	VC
CVV	12.7	***saa***.*ri*	heavy	2	VV
CVVC	9.6	***viet***.*to*	heavy	3	VVC
VC	3.9	***es***.*te*	heavy	2	VC
V	3.9	***o***.*sa*	light	1	V
VV	1.2	***au***.*to*	heavy	2	VV
CVCC	0.6	***kilt***.*ti*	heavy	3	VCC
VVC	0.3	***aal***.*to*	heavy	3	VVC
VCC	0.1	***ark***.*ku*	heavy	3	VCC

It can be seen that the CV syllable type is by far the most frequent one. The syllable types CV, CVC, CVV and CVVC jointly account for 90% of all occurring syllable tokens; all other types are clearly less frequent. The proportion of syllables beginning with a C onset is 91%, and the type C(V)V accounts for 53.1% of all syllable tokens.

There are no frequency counts available on the non-basic syllable types discussed below, but it seems clear that, in most types of discourse, the basic syllable types are overwhelmingly more frequent than the non-basic ones. The non-basic syllable types are given in Table 2:

Table 2. The non-basic syllable types, example words, weight, number of morae and the structure of the rhyme.

Syllable type	Example	Weight	N of morae	Rhyme
(1)				
CVVCC	***Kuort**.ti*	heavy	4	VVCC
(2)				
CCV	***pro**.sent.ti*	light	1	V
CCVC	***pris**.ma*	heavy	2	VC
CCVCC	***prons**.si*	heavy	3	VCC
CCVV	***kruu**.nu*	heavy	2	VV
CCVVC	***staat**.ti.nen*	heavy	3	VVC
(3)				
CCCV	***stra**.te.gi.a*	light	1	V
CCCVC	***stres**.si*	heavy	2	VC
CCCVCC	***sprint**.te.ri*	heavy	3	VCC

In Table 2 the non-basic syllable types have been divided into three groups. Group (1), CVVCC, is very rare, occurring only in some proper names; additional examples are *Suortti, Jotaarkka.* This is the longest, and the only tetramoraic syllable type in the language. The syllable types in Group (2) have a CC onset; theses types are more common than those in the other two groups. The syllable types in Group (3), with a CCC onset, are very marginal (although the words *strategia* and *stress*i are nowadays, unfortunately, very frequent).

There are restrictions on the occurrence of both the basic and the non-basic syllable types related to position in the word. All syllable types are possible as stressed, word-initial syllables, but only some are possible as unstressed syllables later in the word. It will be remembered from above that, in word-internal consonant sequences, the last (or only) consonant is always preceded by a syllable boundary.

8 Canonical word structure, minimal word, minimal utterance

The stems, i.e. uninflected forms, of words are most often disyllabic, and end in an open syllable. Karlsson (2005) reports the lexical frequencies of native, morphologically atomic monosyllabic and disyllabic nouns in the Reverse Dictionary by Tuomi (1972). There were 4958 disyllabic nouns with an open second syllable fulfilling the criteria. Table 3 shows the proportions, in the lexicon, of the eight most frequent disyllabic structures in the material.

Table 3. The proportions of the eight lexically most frequent disyllabic structures in the 4958 nouns in Tuomi (1972) with an open second syllable, as calculated and reported by Karlsson (2005).

Word Structure	Examples	Proportion
CVC.CV	*hihna*, *kukko*, *pentu*	36%
CVV.CV	*jousi*, *laatu*, *nuoli*	19%
CV.CV	*kala*, *peto*, *maku*	15%
CVVC.CV	*haaska*, *juusto*, *lieska*	13%
CVCC.CV	*harppi*, *kalske*, *lamppu*	10%
VC.CV	*ahma*, *olki*, *ämmä*	3%
V.CV	*aho*, *ele*, *äly*	1%
VV.CV	*aamu*, *aika*, *ääni*	1%

The six most frequent structures account for 96% of the native disyllabic nouns ending in an open syllable. Structures not mentioned in Table 3 account for less than 1% each: VVC.CV (e.g. *aalto, aitta, äänne),* VCC.CV (e.g. *ankka, arkki, yrtti),* and X.CVV (e.g. *ehtoo, harmaa, suklaa)*. Karlsson (2005) did not analyse in detail the disyllabic nouns ending in a closed syllable, but reports that their total number was around 800, so about 86% of the disyllabic nouns end in an open syllable. Monomoraic initial syllables account for only 16 percent of the

initial syllables. Thus polymoraic first syllables are clearly favoured, as are C-initial first and later syllables. Karlsson did not analyse trisyllabic and longer nouns in detail but says that “a fast test shows that more than 75% of them too have dimoraic or even heavier first syllables. The same holds across the board for the vocabulary: 75% of the lexemes listed in [Tuomi 1972] have at least a dimoraic first syllable” (Karlsson 2005: 69).

There is no definite upper limit on the number of syllables a word can have, and very long morphologically complex words are possible; for examples see Chapter 1. In describing Finnish sound structure, it may be more interesting to attempt to define the size of the minimal word. Below are listed the shortest syllables and the (numbers of) words consisting of such syllables:

Structure	N of existing words
V	0
VV	3 (*ei* ‘not’ or ‘it does not’, *yö* ‘night’, *ui* ‘(he/she/it) swims’)
CV	7 (*me* ‘we’, *te* ‘you’, *he* ‘they’, *ne* ’they’, *s*e ‘it’, *ja* ‘and’, *jo* ‘already’)
VC	3 (*en* ‘I do not’, *et* ‘you do not’, *on* ‘is’)
CVC	5 (*hän* ‘he, she’, *kun* ‘when’, *jos* ‘if’, *nyt* ‘now’, *kas* ‘look, well, why, lo')
VVC	0

There are no words consisting of just a single vowel, and even the names of vocalic letters and single vowel phonemes are pronounced as phonetically long (phonologically double). Of the words listed above, only *yö* and *ui* are open class words. The words *ei, en* and *et* are forms of the negation verb (*ei* also simply means ‘not’); we do not consider this very special grammatical verb, whose stem consists of just *e*-, to be an open class word. Below are listed some further mono- and disyllabic structures and the number of morphologically atomic nouns having these structures according to Karlsson (2005):

Structure	N of nouns
CVV	24
CVVC	4
VCC	0
CVVCC	0
CV.CV	756
V.CV	61

Listed above are those basic syllable types not mentioned in the preceding list, the rare non-basic type CVVCC and two short disyllabic structures. There are no words of any class with the structures VVC, VCC and CVVCC. Altogether, the number of monosyllabic words is small. The shortest new loanwords have the structure CVV, e.g. *pai* (< Engl. *pie).* Recall from above that words with the structure CVC in the lending language usually become CVC.CV in Finnish, e.g. *pop* > *poppi*.

The shortest unreduced words thus consist of two phoneme segments, their structure is either VV, CV or VC. But if a distinction is made between open class words and closed class words, a different picture emerges. A generalisation can be made that has no native exceptions: the stem of an open class word must contain at least two voiced morae (recall that we do not consider the negation verb, the only apparent exception, to be an open class word). Shortest such open class words are the VV words *ui* 'swim' and *yö* 'night' mentioned above, there are no open class words that contain less than two voiced morae (for example, no VC open class words with a voiced consonant). The next longer open class words with two voiced morae are CVV words like the nouns *puu* 'tree', *kuu* 'moon', *suu* 'mouth', *luu* 'bone'.

Clearly, the productive pattern of word formation includes a condition that new words must contain at least two voiced morae. Why should this be? As will be explained below, the proper phonetic realisation of sentence accent in Finnish presupposes two voiced morae, and open class words are of course often accented, whereas closed class words are accented more seldom.

Apart from open class words in which the requirement is more stringent, a minimal word must thus consist of at least two phonemes. But what is the minimal requirement for an utterance (containing at least one articulated word)? This question has recently been investigated by Suomi (to appear) who elicited short replies from five male speakers to written questions in the laboratory. Here the monomoraic replies are relevant. They included the reduced forms of the replies *on* '(yes, it) is' and *en* 'I do not'. In careful speech these replies are pronounced [on] /on/ and [en] /en/, respectively. But word-final nasals are often deleted in colloquial speech, and the informants were instructed to produce *on* and *en* replies with the final /n/ in some replies, and without it in some other replies; these replies were given to exactly the same set of questions. The replies without the final /n/ are of course monomoraic. Four of the five speakers produced the reduced replies as [ʔoh] and [ʔeh], and the fifth as [oh] and [eh]. The utterance-final aphonematic [h] also occurred, for all speakers, in all other monomoraic utterances (*jo* 'already', *no?* 'well?', and personal pronouns of the form CV). In contrast, the [h] was only sporadically appended to bimoraic replies (such as *joo* 'yeah') by two of the speakers. Suomi interpreted the results regarding the aphonematic [h] as follows. This added, phonologically unmotivated segment is a phonetic mora whose motivation is to guarantee the minimum size of an utterance, viz. two morae. Thus open class words always consist of at least two voiced morae, closed class words must consist of at least two phonemes (of which one must be a mora), and utterances must consist of at least two morae, one of which may be [h], a phonetic mora. Notice that according to this interpretation, the moraic requirement imposed on an utterance is stricter than the moraic requirement of a closed class word. But the requirement on closed class words applies to them as words, and in longer than minimal utterances no other requirement is operative (as in e.g. the utterance *vain he* 'only they', in which no [h] is added to the end). But when a minimal closed class word constitutes an utterance by itself, the requirement concerning a minimal utterance is enacted, and the [h] is added. A similar relationship holds between the minimal syllable and the minimal word: a minimal syllable consists of a single vowel phoneme (which constitutes the syllable nucleus), but a minimal word must contain at least one additional phoneme. Thus just as an utterance must be longer than the minimal word, a word must similarly be longer than the minimal syllable.

The presence of the initial aphonematic [ʔ] in the [ʔoh] and [ʔeh] replies (by four of the five speakers) was verified both auditorily and by acoustic

measurements. The [ʔoh] and [ʔeh] replies differed reliably from all other phonologically vowel-initial replies in terms of mean VOT and the mean intensities of the first five glottal pulses; the glottal pulses had greater intensities in the [ʔ]-initial vowels than in the others, and the differences between the 1st and 2nd glottal pulses and the differences between the 2nd and 5th glottal pulses were also larger in the [ʔ]-initial vowels than in the others.

Recall from above (section 5.3) the observation by Lennes *et al*. (2006) of glottal stops in informal dialogues among speakers from the Helsinki area. Some of the speakers in Suomi (to appear) came from dialect areas in which Itkonen (1964) reported absence of initial glottalisation in dialect interviews. The [ʔoh] and [ʔeh] replies occur only or at least typically in informal speaking situations, and they may be particular instances of the observation by Lennes *et al*. that glottal stops occur in spontaneous, colloquial dialogues.

9 Word-level Prosody

In this chapter we describe word-level prosody, except for the quantity opposition that has already been dealt with above (because the quantity opposition, interpreted as a syntagmatic opposition, is a matter of phonotactics). Many of the empirical findings that we refer to in this chapter are based on our own research, in which the subjects were invariably speakers of Northern Finnish. When speakers are asked to speak out written sentences in the laboratory, they invariably speak a local variety of SSF, not the local dialect. For example, the Northern Finnish speakers in our experiments, from Oulu and its surroundings, do not insert the epenthetic vowel (as in *kolome* for SSF *kolme*) that is a well-known property of their local dialect, nor do they exhibit any other segmental dialectal features. Thus the empirical results do not concern Northern Finnish dialects, but the Northern Finnish variety of SSF. Strictly speaking, then, many of the details of the description may apply to this particular variety of Finnish only. As yet unpublished work by the third author indicates clear durational and tonal differences between Northern Finnish and two southern varieties.

The distinction between word-level and utterance-level prosody is not clear-cut, and some of the phenomena discussed in this Chapter, although concerning words, are influenced by properties of the carrier utterance. For example, the phonetic realisation of accents is clearly dependent on properties of the utterance in which the accented word occurs.

9.1 Degrees of word stress and their phonetic realisation

Three degrees of stress have been traditionally distinguished for Finnish: a syllable may be primarily stressed, secondarily stressed or unstressed. There are empirical phonetic grounds for these three degrees, but not for further degrees. Primary stress is invariably fixed to the word-initial syllable of a word; apparent exceptions will be dispelled below. Secondary stresses occur in longer than trisyllabic words, and they are not fully predictable. Usually, secondary stress falls on the third or the fourth syllable, and later in the word usually on every second syllable except the last one; however, secondary stress may fall on a final heavy syllable if the preceding syllable is light. Ultimately, where secondary stresses fall depends both on the segmental structure of syllables and on morphological structure (for details, see Karlsson 1983: 150–151).

A foot consists of a primarily or secondarily stressed syllable followed by unstressed ones, i.e. feet are left-headed. For example, the word form |ˈ*usko*|ˌ*matto*|ˌ*missa*|ˌ*kaan*| 'not even in incredible (plural form)', in which "|" denotes a foot boundary, contains a primarily stressed syllable and three secondarily stressed syllables, thus altogether four feet, with the last one consisting of a single syllable. The following account of the durational realisation of stress is based on the investigations by Suomi, Toivanen & Ylitalo (2003) and Suomi & Ylitalo (2004) who carefully distinguished between stress and accent (i.e., stress was investigated in target words that were not accented).

As in many other languages, stress is not realised tonally; see e.g. Bruce (1998), Cruttenden (1997), Terken & Hermes (2000). During an unaccented word, F_0 is determined by the adjacent accented words or by boundary tones. Whether there are differences in spectral tilt between vowels in stressed and unstressed syllables, which has been observed for at least English (Huss 1978) and Dutch (Sluijter & van Heuven 1996), is not known. At any rate, as already mentioned above, there is practically no reduction of vowel quality in unstressed syllables, relative to stressed syllables. But stress is realised by variations in segment durations. We next describe the durational realisation of primary stress, and then that of secondary stress.

As a first approximation, it can be said that primary stress is realised by segments having longer durations when they constitute the word's first or second mora, relative to segment durations elsewhere in the first foot. For example, the CV.CV, CVV.CV and CVC.CV words *tuli* 'fire', *tuuli* 'wind' and *tulli* 'customs' can be given as $CM_1.CM_2$, $CM_1M_2.CM_3$ and $CM_1M_2.CM_3$, respectively, where M_n denotes the word's n^{th} mora. Thus in CV.CV words stress is realised by increased duration of both vowels, in CVV.CV by increased duration of the VV, and in CVC.CV words by increased durations of the first-syllable V and C. Thus in *tuli* the second-syllable /i/ (being M_2) has a much longer duration than it has in *tuuli* and *tulli* (where it is M_3). The CVCV(X) word structure is the only one in which stress is realised in both the initial and the second syllable. In fact, the second syllable in words like *pupu* 'bunny' and *koko* 'size', which consist of sequences of two phonemically identical syllables, the duration of the second syllable is longer than that of the initial, stressed one. This difference is largely due to the vowels: the second-syllable single vowel has a much longer duration than that in the first syllable. This is one of the durational alternations that exist in

the first two syllables of words. Detailed discussion of these alternations is postponed to section 9.3. below.

Primary stress is thus invariably associated with the initial syllable. But the largest prominence on a word occasionally falls on the last syllable, as in single-word utterances like *KiiTOS!* 'Thanks!', *NäkeMIIN!* 'Goodbye!', and *AiVAN!* 'Precisely!'. Utterances like these are here interpreted as pragmatically determined instances of unusual accentuation, not as exceptions to the stress rule.

The phonetic realisation of secondary stress has only been investigated to a limited extent. However, some sporadic observations are available. Suomi & Ylitalo (2004) observed that, in the word's second foot, too, the duration of the second-syllable single vowel was longer when it constituted the foot's second mora than when it did not. For example, just as /i/ has a longer duration in *tuli* than in *tulli,* the final /ɑ/ has a longer duration in *tohto|rina* 'as a doctor' than in *tohto|rinna* 'doctor's wife' (where "|" again denotes a foot boundary). However, the difference was smaller than that observed in the first foot, and the authors conclude that the durational correlates of secondary stress are attenuated relative to those of primary stress.

In a way, primary stress is signalled also phonotactically, at least in the majority of words. Of course, only fully native words exhibit the true structural tendencies of a language. Of the ten fully native syllable types, all can occur as initial syllables, and only some of the shorter ones in later syllables. Recall that, according to Karlsson's (2005) computations, 84% of the initial syllables of disyllabic nouns are polymoraic, and that the eight lexically most frequent disyllabic noun structures account for 98% of all disyllabic nouns. That is, the canonical structure of uninflected words has a polymoraic initial syllable and a CV second syllable, CVC.CV being the most common structure, followed by CVV.CV. In other words, in the majority of words, the stressed syllable is heavy while than the unstressed ones are light. As in many other languages, then, stressed syllables tend to be phonotactically more complex than unstressed ones.

But there is one exception, the structure (C)V.CV(X), as in *a.la*, *ka.la*, *va.paa*, *sa.ta.ma*. In this structure, the stressed syllable is light. This is the only word structure in which stress is realised as increased duration of not only of the first syllable vowel but also of the second-syllable vowel. Without this word structure it could be stated that stress is signalled durationally in the first syllable only. But since this word structure exists, it is more parsimonious to state the durational realisation of stress in terms of morae rather than syllables.

Obviously because stress is invariably associated with the word-initial syllable in Finnish, speakers of Finnish have problems with words that deviate from this pattern. Thus e.g. *Sibelius* is pronounced variably, usually ˈ*Sibelius* in the Finnish way, and more seldom as *Si*ˈ*belius* (which is the stress pattern of the name in Swedish, the composer's mother tongue), and similarly with *Finlandia*, *Karelia* etc. In news broadcasts, the stress pattern of a foreign name can change from one sentence to the next, e.g. *Ab*ˈ*bas*, ˈ*Abbas*, and in concert introductions on the radio, the tempo of the next piece can be variably *an*ˈ*dante* or ˈ*andante*. Sometimes the uncertainty of the correct stress pattern results in funny hypercorrections, as when the *Wall Street Journal* is referred to as [woːl striːt ʒurˈnɑl] (the first author has heard this pronunciation several times).

In compound words, which are very common in Finnish, and which usually consist of more than three syllables (because words are usually at least disyllabic), stress assignment usually follows that of non-compound words, e.g. *kesä* 'summer' + *loma* 'holiday' > ˈ*kesä*ˌ*loma* 'summer holiday' is stressed like ˈ*mata*ˌ*lana*, inflected form of ˈ*matala* 'low'. However, a mannerism is spreading in which the second part of a compound is not secondarily stressed but accented. There seem to be two versions of this: either the first part is also accented, or it is unaccented, e.g. either *KESÄLOMA* or *kesäLOMA*. This mannerism, which irritates many, including the present authors, seems to be becoming more and more common on e.g. local radio stations and many commercial TV channels. A similar mannerism concerns the pronunciation of proper names. Usually, a name like Matti Virtanen is pronounced with the first name (Matti) unaccented and the surname accented, as in English. But a mannerism is spreading according to which the first name is accented and the surname is not. This has always been the pronunciation when the first name is contrasted (e.g. *I said MATTI Virtanen, not MIKKO Virtanen*), and similarly the second part of a compound can be contrastively accented (e.g. *Sanoin kesäLOMA, en kesäTYÖ* 'I said summer HOLIDAY, not summer JOB'). But in the new mannerisms no contrast is implied. But here we are anticipating the subject matter of Chapter 10.

9.2 The phonetic realisation of accent

In this section we describe how accents are realised phonetically. Again, it may be emphasised that what follows is based on investigations of Northern Finnish, and does not necessarily apply to other varieties of SSF.

According to our tentative interpretation, three discrete phonological degrees of accentuation are necessary and sufficient in a description of Finnish, namely thematic accent, rhematic accent and contrastive accent (for more details and examples see section 10.2 below). We consider emphatic accent, as e.g. on the last word in the admiring statement *Sinä olet IHANA!* 'You are LOVELY!' to be a gradient, pragmatic variable that does not constitute a phonological degree of accentuation, following Bruce (1998). We have only investigated thematic and contrastive accents, but it is our subjective impression that rhematic accents are, in their tonal realisation, stronger or more salient than thematic ones. There thus appears to be a phonetic hierarchy of the phonological degrees of accentuation such that contrastive accent is phonetically the strongest, rhematic accent is of medium strength, and thematic accent is the weakest. That is, we have investigated, and report below results on, the end points of a hierarchy with supposedly three discrete degrees.

All degrees of accent are realised tonally, which distinguishes accented words tonally from unaccented words: unaccented words have no tonal properties of their own. According to our investigations, only contrastive accent is also realised durationally, whereas thematic accent is not, relative to unaccented words. In the following, therefore, we distinguish between the tonal and durational realisations of accent, in addition to distinguishing thematic and contrastive accent.

By default, accents are associated with the word-initial stressed syllable. There are two kinds of exceptions. Firstly, contrastive accent can be associated with any syllable that carries the contrasted semantic information, as in many other languages. For example, although accent (contrastive or otherwise) is usually associated with the initial syllable of, say, *Helsinki,* contrastive accent can also be associated with e.g. the final syllable in e.g. *Sanoin että tulin HelsinKIIN, en sanonut että tulin HelsinGIStä* 'I said I came TO Helsinki, I did'nt say I came FROM Helsinki'. The other kind of exceptions concern socially important one-word utterances such as *KiiTOS* ('thanks') and *NäkeMIIN* ('goodbye'), already mentioned above. Below, only accents with the default association will be discussed.

Like stress, also accents are mainly realised within the sequence delimited by and including M_1 and M_2. The accentual tune is usually a tonal rise-fall, with the rise taking place during M_1, and a large part of the fall during M_2. Thus in e.g. *tuli* the rise occurs during the first syllable and much the fall during the second one,

while in *tuuli* and *tulli* the rise and a large part of the fall both occur during the first syllable. The F_0 excursions are wider in contrastive than in thematic accent.

Suomi (in preparation) studied in greater detail the segmental anchoring points of the accentual tune. Following the bitonal analysis of intonational phonology originating, in an experimental vein, in the work of Bruce (1977), the tune was defined as the sequence LHL (in which H and L are high and low tones, respectively). Speakers produced the target CVCV, CVVCV and CVCCV words (e.g. *mani* and *nami*, *maani* and *naami*, *manni* and *nammi*) occurring in the carrier sentence *Minun mielestäni* ___ *näyttää paremmalta* 'In my opinion ___ looks better' at three different speaking rates, Slow, Normal and Fast. The target words contained only voiced segments, and segment identities of the words were fully counterbalanced (thus there were also words like *mina* and *nima*, etc.). The speakers produced contrastive, prenuclear accents. The initial L was defined as the temporal location of an F_0 minimum between the beginning of the pre-target syllable and the F_0 maximum within the target, and the H was defined as the temporal location of this maximum. The final L could not be defined on the basis of a local F_0 minimum as the F_0 curve was usually slightly falling well into the next word. Instead, the criterion was the statistical end of the fall between consecutive, structurally defined measurement points.

It was observed that, at each speaking rate and irrespective of word structure, the segmental anchoring points remained the same. The initial L was always anchored to word onset (beginning of the word-initial consonant), the rise beginning at this location. The H was anchored to the end of M_1, and the final L to the vicinity of the middle of the third syllable, which in this material was the first syllable of the word *näyttää*. This last observation is in agreement with that in Suomi (2007), in which the speakers were given no instructions as to speaking rate. Firstly it was observed that the end of the accentual fall was not quite reached by the end of short words like *sei* and *setä*, and that the fall was then completed during the initial syllable of the following word. Secondly, it was observed in longer words like *Seikola(sta)* and *Setälä(stä)* that the fall terminated by the middle of the third syllable. Both studies thus suggest that the final L is anchored to approximately the middle of the third syllable, and that in short words, this syllable in fact belongs to the next word. The results for the normal speaking rate in Suomi (in preparation) are shown in Figure 4 (in which marks representing the three word structures often hide behind each other).

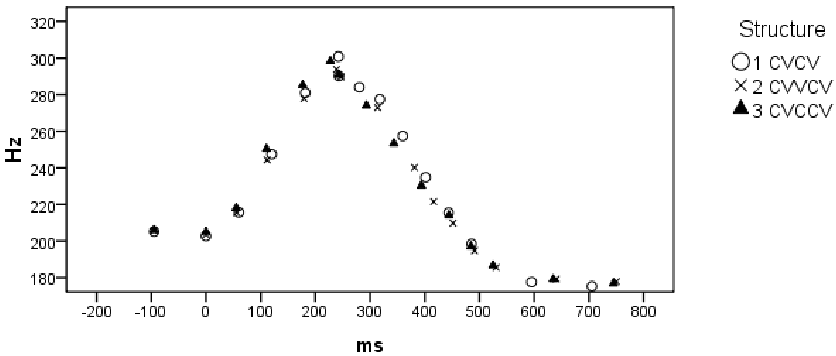


Fig. 4. Mean F_0 values of the normal speaking rate versions of the target words in Suomi (in preparation), from the beginning of the pre-target syllable to the end of the voiced portion of the post-target syllable. Target word onset is at zero ms. For each word structure, the first measurement point (at around -100 ms) was at the beginning of the pre-target syllable, and the last two measurement points (just before and after 600 ms) were at the middle and end of the voiced portion of the post-target syllable, respectively.

Similarly the absolute mean Hz values of the three target tones were the same in each speaking rate, irrespective of word structure. Comparisons across the speaking rates revealed that the speakers varied among themselves in the way speech rate affected the absolute Hz values. Thus for one speaker increasing speaking rate increased the absolute values systematically, for other speakers the effect of rate was not as systematic, and varied from speaker to speaker. Consequently, it seems to be the choice of individual speakers how speaking rate affects the absolute Hz values, and results on this measure therefore depend on the selection of speakers investigated. Corresponding differences have been reported in many other studies, see e.g. Appendix B in Ladd, Faulkner, Faulkner, and Schepman (1999) and the references therein.

These results, which are consistent with our previous experiments in which speaking rate has not been varied, provide strong supporting evidence for bitonal models of tonal phonology. The results are in perfect agreement with the early conclusion by Bruce (1977: 132), who studied the two different lexically determined word accents that occur in (most varieties of) Swedish, that “reaching

a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself. In this way the rise or fall becomes a mere transition, which is necessary in order to go from one level to another", and with the later conclusion by Ladd *et al*. (1999: 1552) that "accent shape has no definition independent of the linguistically specified alignment and F_0 level of the targets of which the accent is composed".

Thus all three tones in the LHL sequence had constant segmental anchoring points and, in each speaking rate, constant target F_0 levels. Moreover, the steep F_0 fall from the H to the final L reached a statistical plateau immediately after the final L, the F_0 curve thus having a concave shape at the final L, see Figure 4 above. For these reasons, Suomi argued that it is indeed appropriate to postulate the tritonal sequence LHL for Finnish (rather than a more usual ditonal sequence).

Apart from the number of tones characterising the accentual tune, the behaviour of the Finnish accentual tune differs from that of the accentual tunes of many other languages. In Finnish, as has been observed in all of our relevant studies, the accentual tune does not vary as a function of word structure. In contrast, the accentual tune is uniform across different word structures (at a given degree of accentuation, in a given speaking rate). In many other languages, the temporal distances between the anchoring points vary as a function of the segmental structure and duration of the accented syllable. For example, Arvaniti, Ladd & Mennen (1998) found both the beginning and the end of rising prenuclear accents in Greek to be anchored to segmental landmarks. The beginning of the rise was temporally aligned with the end of the unstressed syllable preceding the accented syllable, and the end of the rise was aligned with the beginning of the following unstressed vowel. The duration of the F_0 rise varied as a function of the segmental composition of the accented syllable, sometimes vastly. In Finnish, in contrast, *both* the segmental anchoring points *and* the temporal anchoring points are invariant.

Contrastive accent is also realised durationally: certain segment durations are longer in contrastively accented words relative to those in unaccented words and thematically accented words. In the accentual lengthening accompanying contrastive accent, the word-initial consonants as well as M_1 and M_2 are extensively lengthened, and other segments less. Thus in Suomi (2005), in which unaccented and contrastively accented disyllabic CVCV and CVCCV words (e.g. *kana* and *kanta*) were investigated, the accentual lengthening of word-initial consonants was on average 50%, that of M_1 38% and that of M_2 35% (vowels)

and 53% (consonants), and that of other segments 20%. Suomi (2007), investigating up to tetrasyllabic words, observed accentual lengthening to extend from word onset to the end of the third syllable, with minor lengthening appearing on the first segment of the fourth syllable. Word-initial consonants were lengthened on average by 75%, M_1 (excluding the monomoraic personal pronouns, see below) and M_2 both by 58% and other segments by 19%.

Given the constant segmental anchoring points of the accentual LHL tune, and given the quantity opposition, it is inevitable that there are systematic durational alternations across different word structures. As an overview of the situation, Figure 5 shows schematically how thematic accent and contrastive accent are realised in the example words *kana* and *kanta* representing CV.CV and CVC.CV (or $CM_1.CM_2$ and $CM_1M_2.CM_3$) words:

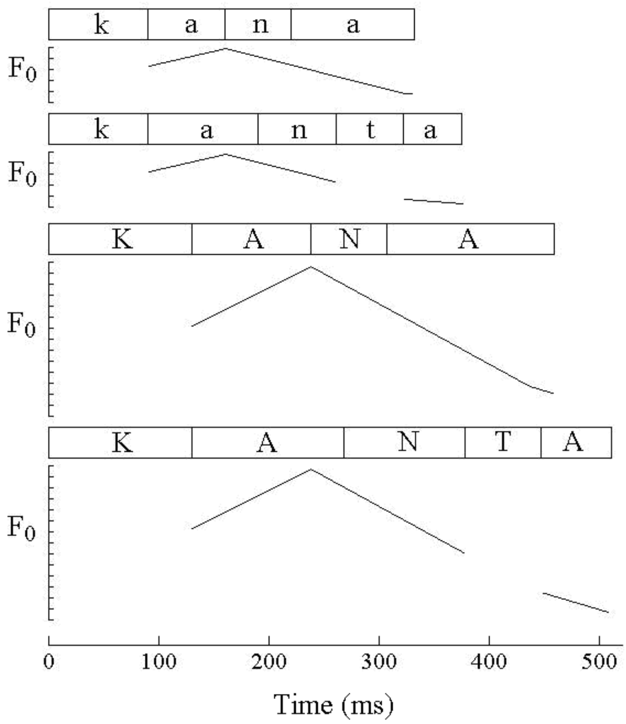


Fig. 5. The realisation of thematic and contrastive accent in the words *kana* and *kanta*. The notation "kana" (lower case) refers to a thematically accented word, the notation "KANA" (upper case) refers to a contrastively accented word.

It can be seen that, in both degrees of accentuation, the duration of the second-syllable /ɑ/ is much longer in *kana* than in *kanta,* and that both the /ɑ/ in the first syllable and the word-medial /n/ have a longer duration in *kanta* than in *kana.* Such durational alternations contribute, in both degrees of accentuation, to the uniformity of the accentual tune. The durational alternations will be discussed in more detail in the next section.

There is one class of words that behaves exceptionally in accentuation. The proper realisation of accent requires two voiced morae, and thus accent cannot be realised in the usual manner in monomoraic words. In Suomi (2007) the monomoraic personal pronouns *he*, *ne* and *se* were among the contrastively accented target words, and these behaved very much differently from the other, longer words, in that the realisations of the accented versions exhibited extensive variation. In the overwhelming majority of tokens of the longer, at least dimoraic words, modal phonation was used throughout the target words, with creaky voice occurring occasionally in syllables following those syllables during which the rise-fall tune was realised. In the monomoraic words, in contrast, there was much more variation, especially as concerns F_0 values. Each of the six speakers produced 10 tokens of contrastively accented monomoraic words. One speaker realised a rise-fall in all 10 tokens, with strong and long [h]-like aperiodic noise always following the voiced portion of the vowel. A second speaker realised eight high level tunes and two rises, with no aperiodic portion at the end of the vowel in any token, and a third speaker realised seven rises, two high level tunes and one fall, with a pause after the target word in three tokens (but the speaker did not pause after the longer target words). The other speakers realised combinations of these characteristics: variable F_0 tunes, aperiodic noise following the voiced portion of the vowel, or a pause following the target word. All of these realisations sounded familiar and perfectly natural. The modally phonated vowels followed by aperiodic noise did not give an impression of double vowels; rather, the noise just sounded like a way of indicating that the word is contrastively accented, and to perform the same function as the pauses after the target word. The variability is an indication that monomoraic words are metrically marginal; recall from Chapter 8 that the stem of an open class word must contain at least two voiced morae.

9.3 Segment durations and moraic structure

Differences in intrinsic duration make it difficult to systematically study structural factors that affect segment durations using real word materials, as there are many accidental gaps in the vocabulary. For example, Lehtonen (1970) had to compare two word structures at a time, e.g. CV.CV and CV.CVV, using real words that, in both word structures, contained qualitatively identical segments and only differed with respect to quantity (e.g. *sata* and *sataa).* In this comparison, certain CV.CV and CV.CVV words were used. For other comparisons, e.g. that of the word structures CV.CV and CVV.CV, Lehtonen had to use another set of CV.CV words. For such reasons, it was not possible to directly compare segment durations in e.g. CV.CVV and CVV.CV structures. Consequently, we can only occasionally refer to Lehtonen's findings; moreover, Lehtonen did not investigate accentuation and how it relates to segment durations, which has been the focus of our research. In this section, we mostly discuss two studies in which such problems have been evaded by using fully counterbalanced nonsense materials. The studies are Suomi & Ylitalo (2004) and Suomi (submitted); Suomi & Ylitalo did not consider their data or report results from the present perspective. The results presented below have been reported in more detail, including the results of statistical tests, in Suomi (2006).

In Suomi & Ylitalo (2004), segmentally fully balanced nonsense items representing the trisyllabic, one-foot word structures CV.CV.CV, CV.CVC.CV, CV.CVV.CV, CV.CVV.CVV, CVC.CV.CV, CVC.CVC.CV, CVV.CV.CVV and CVV.CVV.CVV were investigated. Every structure was represented by 18 segmentally different items. The consonant was always /p/, /t̪/ or /m/, and the vowel one of /i/, /ɑ/ or /u/, both the vowels and the consonants occurring phonologically single or double. All items contained only one occurrence of each phonetic consonant and vowel, and the segmental phonetic composition was counterbalanced across the three consecutive syllables. For the CV.CV.CV structure, the 18 items are shown in Table 4.

The items were embedded in a constant frame sentence, in which they were at most weakly accented. New analyses of these data have now been performed, with a view on moraic structure and segment durations. Let us first look at vowel durations.

Table 4. The 18 items of the CV.CV.CV structure in Suomi & Ylitalo (2004). The items representing the other structures were constructed from these by doubling consonants and vowels, as appropriate (for example, to obtain *mippattu* and *miipaatuu*).

mipatu	*matupi*	*mupita*	*pitamu*	*pamuti*	*putima*
timapu	*tapumi*	*tumipa*	*mitupa*	*mapitu*	*mutapi*
pimuta	*patimu*	*pumati*	*tipuma*	*tamipu*	*tupami*

The new analyses revealed four statistically distinct, non-contrastive and complementary duration degrees for single vowels, and similarly three such degrees for double vowels. The duration degrees were obtained by grouping the vowels according to their moraic affiliation; recall that e.g. CVV.CVV.CVV is moraically $CM_1M_2.CM_3M_4.CM_5M_6$ and CVC.CVC.CV is moraically $CM_1M_2.CM_3M_4.CM_5$. It was observed that the durations of single vowels constituting the word's third or later mora did not differ from each other statistically. Taking this into account, the results (together with those from another experiment discussed below) are shown in Table 5.

Table 5. The mean durations (in ms) of the vowel classes $V_{(1)}$ – $V_{(4)}$ and $VV_{(1)}$ – $VV_{(3)}$ in Suomi & Ylitalo (2004) and Suomi (submitted) (columns S & Y and S, respectively), the duration of (V)V divided by the duration of the [very short] degree, in both materials (again, columns S & Y and S, respectively), the labels of the duration degrees, and example word structures (where "X" denotes any phonotactically possible additional segments).

	S & Y	S	S & Y	S	Label of duration grade	Example structure(s)
$V_{(1)}$	48	75	1.0	1.0	[very short]	CVV.C$\underline{V}$X, CVC.C$\underline{V}$X
$V_{(2)}$	58	104	1.2	1.4	[short]	C$\underline{V}$.CVX
$V_{(3)}$	73	126	1.5	1.7	[longish]	C$\underline{V}$C.CVX
$V_{(4)}$	84	158	1.8	2.1	[long]	CV.C$\underline{V}$X
$VV_{(1)}$	135	-	2.8	-	[very long]	CVV.C$\underline{VV}$.C$\underline{VV}$
$VV_{(2)}$	142	-	3.0	-	[long] + [very short]	CV.C$\underline{VV}$X
$VV_{(3)}$	149	-	3.1	-	[longish] + [longish]	C$\underline{VV}$.CX

As will be explained below, the double vowel categories $VV_{(2)}$ and $VV_{(3)}$ can be interpreted as sequences of two single vowel duration degrees.

Suomi (submitted) also studied segment durations using fully counter-balanced nonsense materials. Triplets of items were constructed, with one member of a triplet having the structure $C_1V_1.C_2V_2$, and the other two members the structure $C_1V_1C_2.C_3V_2$. In each member of a given triplet, C_2 was either /t̪/, /m/, or /l/. In one member with the structure $C_1V_1C_2.C_3V_2$ the consonant in the C_3 position was /s/, in the other member with the same structure C_3 was /n/. Example triplets, with V_1 and V_2 always /ɑ/ and C_1 always /p/, are *pata, patsa, patna; pama, pamsa, pamna* and *pala, palsa, palna.* The nine items in these three triplets jointly illustrate the experimental design. In this material, the second-syllable vowel in CVC.C<u>V</u> items constitutes M_3, the first-syllable vowel in C<u>V</u>.CV items constitutes M_1 not immediately followed by M_2, the first-syllable vowel in C<u>V</u>C.CV items constitutes M_1 immediately followed by M_2, and the second-syllable vowel in CV.C<u>V</u> items constitutes M_2 not immediately preceded by M_1. The words were embedded in a constant frame sentence, in which they were contrastively accented. Again, four statistically distinct duration degrees for single vowels were observed (see Table 5).

The absolute durations of the respective duration degrees in Table 5 are systematically longer in Suomi (submitted) than those in Suomi & Ylitalo (2004). Apart from possible differences in mean speaking rate, this difference is likely to be due to the fact that the target words were contrastively accented in the former but unaccented in the latter. Accentual lengthening has been observed to influence all segments to the end of the third syllable (Suomi, 2007), and thus even the [very short] vowels in Suomi (submitted) were very probably lengthened, relative to unaccented words. However, accentual lengthening lengthens especially M_1 and M_2, i.e. the [short], [longish] and [long] vowels, and this seems to explain why the ratio of these duration degrees to the [very short] degree is larger in Suomi (submitted) than in Suomi & Ylitalo (see the 3rd and 4th columns in Table 5). Instead, the ratios [longish]/[short] are much more similar in the two materials, namely 73/58 = 1.26 in Suomi & Ylitalo and 126/104 = 1.21 in Suomi (submitted), and similarly with the ratios [long]/[short]: 84/58 = 1.45 and 158/104 = 1.52, respectively. That is, accentual lengthening does not have much influence on the relative durations among the [short], [longish] and [long] duration degrees, but it increases the relative difference between these degrees and the [very short] one.

Thus the four duration degrees for single vowels have been observed in both unaccented words and contrastively accented words. Accentual lengthening does not destroy the durational pattern observable in non-accented words, it just makes the individual duration degrees longer, and increases the difference between the [very short] degree and the longer ones. It has been known for long that three such duration degrees exist, namely the degrees denoted as [very short] and [long] here, and an intermediate degree. But that there are in fact two intermediate degrees is not a novel observation in varieties of Finnish. The degrees corresponding to our [short] and [longish] degrees have also been observed in Central Finnish by Lehtonen (1970). In the relevant comparisons, between the structures CV.CV and CVC.CV, between CV.CVC and CVC.CVC, between CV.CVV and CVC.CVV, and between CV.CV.CV and CVC.CV.CV, the pairwise compared words representing the structures were always segmentally identical except for the quantity of the medial consonant, e.g. *laki, lakki*, but across the four pairwise comparisons both vowel and consonant identities may have varied and may have influenced the mean durations, and we therefore present no numerical data. In each compared structural pair, the first-syllable vowel corresponds to our $V_{(2)}$ in the first member of the pair, and to our $V_{(3)}$ in the second member. In each comparison, Lehtonen reports that $V_{(2)}$ had a reliably shorter duration than $V_{(3)}$. As was shown by Suomi (2006), the duration degrees $V_{(2)}$ ([short]) and $V_{(3)}$ ([longish]) can also be observed in Suomi (2005), in both unaccented and contrastively accented versions of the target words. In brief, the existence of four duration degrees for single vowels has been observed in four studies, with both real words and nonsense materials, under two degrees of accentuation, and in two varieties of Finnish (given that the extreme degrees have been well established long ago).

As concerns the double vowels in Suomi & Ylitalo (2004), the authors reported that in the structure CVV.CVV.CVV the sequence VV had a significantly longer duration in the first syllable than in the later syllables, and that in the structure CVV.CV.CVV the first syllable VV had a significantly longer duration than the VV in the third syllable; these observations reflect the lengthening effect of stress on the words' first two morae. The authors did not explicitly compare segment durations in the second and third syllables, but Suomi (2006) performed such comparisons in the same material and found that in the structure CVV.CVV.CVV — the only structure in which a controlled comparison was possible — there was no statistical difference between the second-syllable VV

and third-syllable VV. Thus there are grounds for distinguishing, in Table 5, between $VV_{(1)}$ (VV in second and third syllable) and $VV_{(3)}$ (VV in the first syllable). Suomi (2006) also established that the observed duration of $VV_{(2)}$, 142 ms, was not statistically distinct from the sum $V_{(4)} + V_{(1)}$, 132 ms, and thus there are numerical grounds for interpreting $VV_{(2)}$ as a combination of $V_{(4)}$ and $V_{(1)}$. The observed duration of $VV_{(3)}$ in turn, 149 ms, is very close to twice the duration of $V_{(3)}$, 146 ms, and thus there are numerical grounds for interpreting $VV_{(3)}$ as a sequence of two $V_{(3)}$'s. Below, it will be shown that these interpretations of VV in CVC<u>VV</u> and C<u>VV</u>CV as sequences of particular V categories result in very simple distributional rules of the duration degrees. In contrast, the duration of $VV_{(1)}$ cannot be meaningfully interpreted to be the sum of the durations of two single vowel categories. The rules stating the distribution of the vowel duration degrees will be given below, after a discussion of consonant durations.

Consonant durations have been examined to a considerably lesser extent than vowel durations; Lehtonen (1970) studied consonant durations extensively, but since real words were used, word structures can be reliably compared only pairwise. Nevertheless, the materials of Suomi & Ylitalo (2004) contain single and double consonants in a number of structural positions. The single consonants invariably constituted the syllable onset and were thus non-moraic. Suomi (2006) grouped them with respect to two variables: position in the word (at the beginning of the first, second or third syllable) and the following context (a single or a double vowel). Table 6 shows the measured grand mean durations as a function of these groupings.

The mean durations of all consonant groups with a different duration degree label differed from each other significantly. Thus a consonant had a longer duration word-initially than when initial in the second or in the third syllable, but there was no difference between the latter two positions. We are not sure how this finding should be interpreted. On the one hand, the lengthening could be attributable to stress; Gordon (1997) observed, in Estonian, longer durations of nasals in the onset position of stressed syllables than in the onset position of unstressed syllables. On the other hand, the lengthening is similar to the word-initial lengthening in English (White, 2002) where e.g. /p/ has a longer duration in *porter* than in *report* (in both words /p/ occurs as onset of a stressed syllable). In Finnish, of course, an effect of the word-initial position cannot be dissociated from an effect due to the onset position of a stressed syllable.

Table 6. The mean durations of the single and double consonants in Suomi & Ylitalo (2004) in the contexts __V and __VV, labels of the duration degrees and example word structures. $C_{(1)}$ = onset C in the first syllable, $C_{(2)}$ = onset C in the second syllable, $C_{(3)}$ = onset C in the third syllable, $CC_{(2)}$ = the first C is M_2, $CC_{(3+)}$ = the first C is the word's third or later mora. The differences between $C_{(2)}$ and $C_{(3)}$ did not reach significance in the context __V nor in the context __VV (but the differences between the contexts were always significant), hence both $C_{(2)}$ and $C_{(3)}$ are labelled [short].

	__V	__VV	Label of duration grade	Example structures
$C_{(1)}$	92	99	[longish]	CV.CX, CVV.CX
$C_{(2)}$	80	94	[short]	CV(V).CVCX, CV(V).CVVX
$C_{(3)}$	73	89	[short]	CV(V).CV(V).CV, CV(V).CV(V).CVV
$CC_{(2)}$	146	-	[very long]	CVC.CVCV, CVC.CVC.CV
$CC_{(3+)}$	116	-	[long]	CV.CVC.CV, CVC.CVC.CV

As concerns the effect of the following vocalic context, the duration of C was on average 12 ms longer before VV than before V. Suomi & Ylitalo (2004) computed that in Lehtonen (1970) the corresponding reliable difference in five structural pairs compared was on average 13 ms. What causes this small but systematic difference is unclear; several mutually exclusive explanations have been offered, but they are not convincing. At any rate the lengthening of C before VV is clearly a phenomenon different from the duration grades proper that are distinguished here. The duration degrees proper occur in mutually exclusive environments, i.e. they are complementary, whereas the following vocalic context seems to only modify duration degrees.

It can be computed from Table 6 above that the mean duration of C before a single vowel in the onset position of syllables other than the word-initial one was (80 + 73)/2 = 77 ms, which can be labelled as a [short] duration of C in the material. The mean duration of a word-initial C before a single vowel (92 ms) in turn can be labelled a [longish] duration of C.

As concerns double consonants in Suomi & Ylitalo (2004), the only grouping variable available was position in the word; unlike single consonants, double consonants only occurred before a single vowel. CC straddled either the boundary of the first and second syllable in which case the first C was M_2, or the boundary of the second and third syllable in which case the first C was the 3rd or a later mora. As can be seen in Table 6 above, CC had a longer duration in the former situation ($CC_{(2)}$) than in the latter ($CC_{(3+)}$). It is of course not possible to directly measure how the total duration of a double consonant straddling a syllable boundary is distributed to the two syllables (as a double consonant is phonetically a single segment), but in view of the situation obtaining in sequences of two different consonants in the same structural position (see below), we infer that since the first segment of the CC occurring at the boundary of the first and second syllable constitutes M_2, this accounts for the longer duration of $CC_{(2)}$ over $CC_{(3+)}$. That is, since the first consonant in a CC sequence consisting of two qualitatively different consonants tends to have a longer duration when it constitutes M_2 than otherwise (see below), we infer that the first segment in a double CC consonant similarly tends to have a longer duration when it constitutes M_2, although, in this case, this cannot be directly measured.

We next look at consonant durations in word-internal sequences of two different consonants in Suomi (submitted); in all CC sequences to be discussed, the first consonant in the sequence constitutes M_2. The nonsense experimental materials were explained above, but let us repeat that a representative series of items is *pata, patsa, patna; pama, pamsa, pamna* and *pala, palsa, palna.* It was observed that in the structure CVC_2V (e.g. *pata),* in which C_2 is not a mora, the mean durations of /t̪/, /m/, and /l/ were significantly different from each other (99 ms, 80 ms and 56 ms, respectively). But in the structure CVC_2C_3V (e.g. *patsa),* in which C_2 is M_2, the durations of /t̪/, /m/, and /l/ were longer throughout (on average 124 ms), and apart from a couple of exceptions with an independent explanation (elastic compensation within the word), statistically equal. But the duration of C_2 is not always longer in the structure CVC_2C_3V than in the structure CV_1C_2V.

The durational behaviour of a consonant constituting M_2 can be summarised as follows. Inherently short consonants (voiced resonants) always undergo second-mora lengthening when they constitute M_2. Inherently long consonants (voiceless obstruents) are lengthened in contrastively accented words, but they are not lengthened (or may even be shortened) in unaccented or thematically accented

words. For more details see Suomi (2006), where results from relevant studies were analysed from this perspective.

Let us now summarise the above findings of vowel and consonant duration alternations by presenting the rules that state the distributions of the duration degrees. For the vowels, two rules are needed, a V-rule and a VV-rule. By convention, the VV-rule is to be applied to double vowels, and if the rule is not applicable, then the V-rule is to be applied separately to both of the V's in the VV sequence; the V-rule is of course applied to all single vowels. The rules, with example word structures, are as follows:

VV	→	[very long] if it does not contain M_2	CVS.CVV.CVV
V	→	[very short] if it is not M_1 or M_2	CVS.CV.CVX, CV.CVV
	→	[short] if it is M_1 not immediately followed by M_2	CV.CVX
	→	[longish] if it is contained in the sequence M_1M_2	CVC.CVX, CVV.CX
	→	[long] if it is M_2 not immediately preceded by M_1	CV.CVCX, CV.CVV

Notice that the rules do not invoke the syllable, they only refer to moraic structure (with reference to word onset). We venture to claim that these rules, given the specification of how the VV-rule is to be applied, correctly capture all of the observed vowel duration degrees in the first, primarily stressed foot, with one exception: we have not studied the word structure (C)V.VX, in which there is a syllable boundary between the word's first two vocalic morae. In such a sequence, it might be very difficult to reliably segment the vowel sequence to two phonetic vowels.

Thus, a double vowel is [very long] according to the VV-rule in e.g. the structures CVV.CVV.CVV and CVC.CVV (as indicated by underlining). But the VV-rule is not applicable to double vowels in the first syllable in CVV.CX words

because these contain M_2, nor to the second-syllable VV in CV.C<u>VV</u> words, because this VV sequence also contains M_2. Instead, the V-rule must be applied. In C<u>VV</u>.CX words both consecutive V's are [longish] because they are contained in the sequence M_1M_2, and in CV.C<u>VV</u> words the first V in the second-syllable VV sequence is M_2, it is [long], and the second V, being M_3, is [very short].

Thus in those syllables of the first foot that contain M_1 or M_2, the segments during which accent is mainly realised tonally, complex but lawful alternations between [short], [longish] and [long] vowel duration degrees occur. But in later syllables not containing M_1 or M_2, there appear to be no subphonemic durational alternations. Thus, in such later syllables, single vowels are always [very short], double vowels always [very long]. These can be characterised as default durations: vowels have such durations when the durations only serve to signal the quantity opposition. Looking at Table 5 above, it can be seen that, in the materials in Suomi & Ylitalo (2004), the VV/V duration ratio in these later syllables was 135/48 ms, i.e. the duration of VV was 2.8 times that of V.

Despite the existence of four degrees of single vowel duration, there is thus usually a safe durational margin between single and double vowel durations, given constant speaking rate. For example, it can be computed that in Suomi & Ylitalo (2004), the first-syllable VV/V duration ratio in C<u>VV</u>CVX and C<u>V</u>CVX words was 149/58 = 2.6. The [long] second-syllable vowel in CVC<u>V</u> words, which is traditionally called the half-long vowel, is the single vowel most prone to being confused with a double vowel in the same position. In the same material, the second-syllable VV/V duration ratio in CVCVVX and CVCVX words was considerably smaller than in the other positions, 142/84 = 1.7. Perhaps not surprisingly, uninflected native CVCVV words are remarkably rare, and Karlsson (2005) reports no such native morphologically atomic noun (see Chapter 8), although there are loanwords like *filee*, *revyy* and a few place names like *Akaa*, *Lepaa*, *Sipoo*. In the first author's dialect, single and double vowels do not contrast in non-initial syllables (i.e., CVCV and CVCVV words are homophonous).

The distribution of consonant duration degrees can be described, tentatively, with two rules, the C-rule and the CC-rule (applicable to single and double consonants, respectively):

C → [longish] in the context #__ or if it constitutes M_2

→ [short] elsewhere

CC → [very long] if it contains M_2 (with certain exceptions)

→ [long] elsewhere

Since a VV in the first syllable is analysable as the sequence [longish] + [longish] and since a C that constitutes M_2 can also be characterised as [longish] (i.e., consonants in this position tend to have constant duration irrespective of inherent duration, a duration longer than that of inherently short consonants elsewhere), the generalisation can be made that, if the first syllable contains M_1 and M_2, both of them are [longish].

It was computed above that the mean duration of C before a single vowel in the onset position of syllables other than the word-initial one was 77 ms, so the CC/C duration ratio later in the word was 116/77 ms, i.e. the duration of CC was 1.5 times that of C (compared to the corresponding ratio 2.8 for vowels). Thus, late in the primary-stressed foot at least, the vowel quantity opposition seems to be signalled more efficiently than the consonant quantity opposition.

It is noteworthy that the duration degrees occurring in the first two syllables of words are fully determined by the choice of the duration degree of a word's first mora. Let us show this for the structures CV.CV(CX), CVS.CV(CX) and CVV.CV(CX):

CV.CV $C_{[longish]}V_{[short]}.C_{[short]}V_{[long]}$

CVS.CV $C_{[longish]}V_{[longish]}C_{[longish]}.C_{[short]}V_{[very\ short]}$

CVV.CV $C_{[longish]}V_{[longish]}V_{[longish]}.C_{[short]}V_{[very\ short]}$.

Notice that the sequence of duration degrees in the latter two structures is the same, reflecting the metrical identity of these structures. Thus there are systematic durational alternations in vowels that constitute M_1 and M_2, segments during which most of the accentual tune is realised, but no such alternations in vowels occurring later in the word (M_3, M_4 and M_5). The alternations make it possible for the anchoring points of the LHL accentual tune to be the same irrespective of word structure, and for the temporal distances between the anchoring points to be the same irrespective of word structure. Especially important would seem to be the constant anchoring of the H tone at the end of the word's first mora, as what immediately follows the first mora is important to the quantity oppositions, as will be argued in more detail below. Consequently, we wish to claim, the distributional rules of the duration degrees not only correctly state the distributions, they also explain why the distributions are what they are.

As has been seen, the durational alternations discussed above are there also in unaccented words, in which the alternations have no tonal motivation. There seems to be a dichotomy between contrastive accent on the one hand, and the other degrees of prominence, on the other: accentual lengthening in the former, shorter segment durations in the latter, with no difference between unaccented and thematically accented words. In his review of Suomi (submitted) Laurence White pointed out that the observation of four degrees of vowel duration in unaccented words in Suomi & Ylitalo (2004) suggests that, in Finnish, structural influences bear on duration even in the absence of prosodic timing effects (such as domain-edge or domain-head lengthening, see section 9.4 below), and that this contrasts with his model as applied to English, in which compensatory processes would only be observed within prosodically-lengthened constituents. White further suggested that this may relate to the difference in tonal alignment patterns between English and Finnish: in English, the slope and duration of F_0 excursions are variable, while the alignment points of the maxima and minima appear fixed; in Finnish, however, the shape and duration of the contour, as well as the alignment points, are fixed. It may be, White concluded, that preliminary durational adjustments are necessary even in the absence of accent to allow this uniformity. We fully agree with this suggestion.

The above rules stating the distribution of the vowel and consonant duration degrees were based on experiments in which several different vowel phonemes represented vowels as a class, and similarly for consonants. Suppose the rules were used to determine segment durations in synthetic speech. The absolute

duration values in Tables 5 and 6 could be taken as a starting point. But for optimal results, these values would have to be modulated by several factors, e.g. variations in intended speaking rate, degree of accentuation, inherent durations of segments, and their segmental contexts. Suomi (submitted) explicitly studied the latter two factors in the *pata, patsa, patna; pama, pamsa, pamna* and *pala, palsa, palna* word types. It was observed, for example, that t̪/, /m/, and /l/ exhibited their distinct inherent durations in some contexts but not in others (see above), and that the observed distinct durations were almost always compensated by the duration of some other segment in the same word. In other words, the segments within a word exhibited elastic behaviour. The rather complex results were summarised by postulating the following bipartite timing principle:

1. Lengthen any segment that constitutes M_1 or M_2, relative to the segment's other positions.
2. Otherwise compensate, wholly or partially, any durational differences in one segment position by inverse differences in another segment position (with the goal that all CVCV words have equal durations, all CVCCV words similarly have equal durations, and differences in the total durations of CVCV and CVCCV words are minimised).

Almost always, the results were consistent with this principle. The point here is that an exhaustive segment duration algorithm must include information on inherent durations and on contextual elasticity.

Thus several major factors are known to affect segment durations in Finnish words in utterance positions in which durations are not likely to be affected by the boundaries of higher-level prosodic units. Apart from inherent durations specific to each segment, and apart from variations in speech tempo, factors that presumably affect segment durations in all languages (even if not in exactly the same way), the most conspicuous language-specific factor affecting vowels and consonants as segment classes is the quantity opposition: in a given position, a double segment always has a longer duration than a single segment. The second language-specific factor is moraic structure that affects especially vowel duration: a given vowel phoneme has a different duration depending on whether it constitutes M_1, M_2 or a later mora, and on whether M_1 and M_2 are contiguous or not. These durational alternations are present irrespective of a word's degree of prominence. A third factor is accentual lengthening accompanying contrastive accent: word-initial consonants as well as M_1 and M_2 are lengthened much more

than other segments within the durational domain of accentual lengthening. And there is interaction between these three factors.

Consequently, e.g. the final /ɑ/ in *laama* may have a very short duration because, by virtue of being M_3, it has the [very short] duration degree, and because the carrier word is unaccented, whereas the final /ɑ/ in contrastively accented *LAMA* may have a very long duration because, by virtue of being M_2 that is not immediately preceded by M_1, it has the [long] duration degree, and because the carrier word is contrastively accented. In Suomi *et al*. (2003), the mean duration of the final single vowel was 53 ms in *laama* type unaccented words, and 137 ms in *LAMA* type contrastively accented words, i.e. 2.6 times longer duration in the latter words. In the *laama* type the mean duration of the first-syllable double vowel varied between 144 ms (thematic accent) and 198 (contrastive accent); there was thus hardly any difference between the durations of the final /ɑ/ in *LAMA* (137 ms) and the first syllable /ɑɑ/ in *laama* (144 ms). That is, even in materials spoken with a constant tempo, the duration of a single vowel may vary extensively, and under certain conditions be practically equal to that of a double vowel under certain other conditions. But this variability is not chaotic, even though all conditioning factors are certainly not yet known.

Less is known about consonant durations, but differences in inherent durations are observable in at least the non-moraic word-medial position in unaccented CV$\underline{C}$V words, but these differences disappear in at least contrastively accented CV$\underline{C}$CV words. In the latter M_2 position, inherently short consonants are always lengthened; inherently long consonants are lengthened in contrastively accented words but are not lengthened in unaccented words, and may even be shortened. Otherwise, single consonants have a longer-than-elsewhere duration word-initially and before VV. Double consonants have a longer duration when the first consonant constitutes M_2 than when it does not; this does not necessarily apply to inherently long consonants.

Let us return to the durational and tonal interplay in the realisation of accent. The constant duration of the accentual tune in Northern Finnish, irrespective of word structure, is achieved by the durational alternations discussed above. But *why* should the language aim at such uniformity? It may seem paradoxical that, in a full-fledged quantity language like Finnish, segment durations nevertheless vary extensively but the accentual LHL tune is highly constant, while in many non-quantity languages, in contrast, the number of segments (and hence the duration) of the accented syllable determines the tonal realisation. Superficially at least, one

would rather expect that, in a quantity language, segment durations vary only (or predominantly) to signal the quantity opposition, and that the accentual tune is varied accordingly.

But perhaps there is no paradox. In Suomi (in preparation), variation in speaking rate had no effect on the relative segment durations (although absolute durations of course varied). Perhaps the temporal and tonal constancy, a rate-dependent clock as it were, provides a frame of reference against which it is easier to perceive the important quantity distinctions than would be in a system in which temporal distances between target tones vary as a function of segmental material. Perhaps the clock, when rate is varied, does not allow relative segment durations to change? Consider Figure 6 from Suomi (in preparation); for explanation of the experimental design see section 9.2 above.

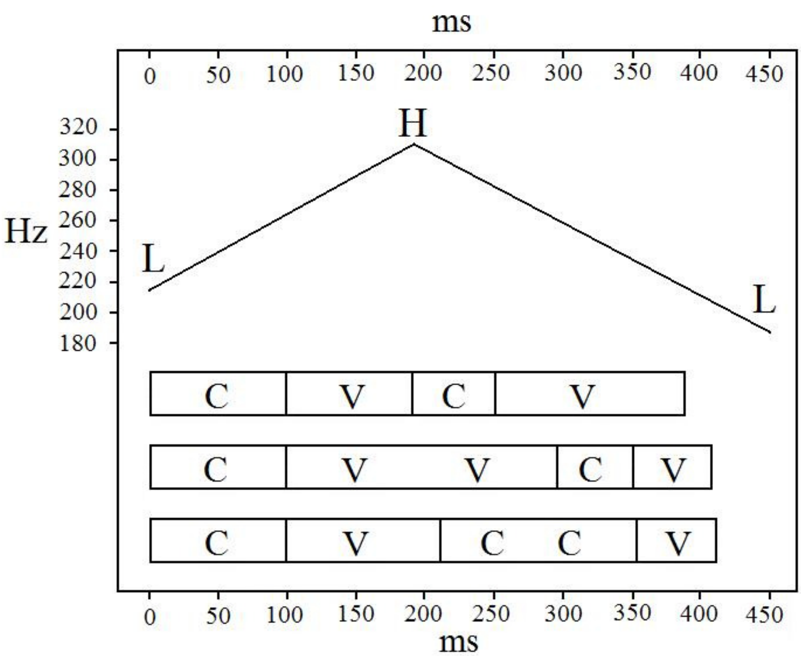


Fig. 6. The relationship of the tonal LHL accentual tune (in Hz, schematised with respect to the intermediate values between the target tones) and the segment durations (in ms) in the normal speaking rate in each of the three word structures investigated in Suomi (in preparation).

Recall from above that the initial L was anchored to word onset, the H to the end of M_1, and the final L to the vicinity of the middle of the third syllable as counted from word onset (i.e., to the vicinity of the middle of the first syllable of the next word). And notice in Figure 6 that the very long duration of the second-syllable V in the CVCV words contributes to the circumstance that the H-to-L distance is invariant across word structures. Suomi (in preparation) argued as follows. When the listener reaches the accentual H, she can deduce that she has just heard the word's first vowel phoneme (M_1). Assuming that the word begins with a consonant, as in the materials investigated, the listener can deduce that she has heard a word-initial CV sequence. If the vowel continues after the H, and if F_0 falls considerably during this continuation, the first syllable is very likely to have the structure CVV. But if, instead, a consonant starts at the H, and if a large part of the F_0 fall occurs during this consonant (if it is voiced) and there is only a low continuation of the fall during the vowel following the consonant, then the listener can deduce that the consonant is very probably a double one, and that she has so far heard the structure CVCCV. (If, instead, there are two consecutive qualitatively different consonants after M_1, then it is all the clearer that the structure must be CVCCV.) But if a consonant starts at the H and is followed by a vowel during which most of the fall occurs, starting at a relatively high F_0 level and reaching a very low level, then the listener can deduce that she has very probably heard a word of the structure CVCV. Thus, many quantity judgements could be made on the basis of rather crude tonal cues alone. Essential to these tonally deduced probabilities is the constant anchoring of the H tone to the end of M_1. If the H were anchored to the end of the first syllable instead, the probabilities would fail. But as it is, the H signals to the listener that what immediately follows is highly relevant to quantity judgements. We are not claiming that these tonal cues are always available in accented words; we have observed variations from the rise-fall tune (including occasional falling tones), but the rise-fall pattern has always emerged as the dominant, average pattern.

Consider a conceivable alternative system, using the example words *tuli* 'fire' (CVCV) and *tuuli* 'wind' (CVVCV). Imagine that the accentual LHL tune were such that there is always a rise during the first syllable, and a fall during the second syllable. In such imaginary Finnish, syllable structure would have the sort of effect on the duration of the rise it has in e.g. Greek and British English prenuclear accents, and the anchoring point of the H would give no clue about the quantity of the first syllable vowel (as the anchoring point would always be the

end of the vowel). Imagine further that the first-syllable V – VV opposition had no repercussions elsewhere in the word. In such a language, it might not be easy to maintain the quantity oppositions. In the imaginary Finnish, the words *tuli* and *tuuli* would differ from each other only with respect to properties of the initial syllable.

It is not improbable that the imaginary system just sketched, in which perceptual cues of quantity distinctions would be strictly local, would be much more vulnerable to perceptual confusions than the really existing system, in which more global durational and tonal cues signal quantity distinctions. As things are in Finnish, the differences between the example words *tuli* and *tuuli* are signalled, in addition to the tonal cues discussed above (present in accented words), by a robust durational difference in the second syllable: a much shorter duration of the second-syllable vowel in *tuuli* than in *tuli*.

Several authors have suggested that tonal cues may help Finnish listeners to perceive the quantity oppositions in accented words, see e.g. O'Dell (2003: 77) and the references therein. In the framework of his oscillation model of speech timing, O'Dell says (on p. 77–78) that "it may be conjectured that pitch movement helps to provide the listener with an indication of the speaker's 'time line' (ie. helps synchronize listener and speaker) for the purpose of making quantity judgments", and notes that the results of his perception experiment, in which spectral, intensity and fundamental frequency differences extracted from natural speech tokens of *tuli* and *tuuli* words were synthesised and systematically varied and their effects on the perception of the quantity opposition were studied, were consistent with such a hypothesis. The results of Järvikivi, Aalto, Aulanko & Vainio (2007) are also consistent with this hypothesis. The authors manipulated the durations of the vowels of original CVCV words (e.g. *sika*) stepwise so that, at the other extreme, the durations were appropriate to CVVCV words (e.g. *siika*), and the tone of first syllable vocalic portion had two different realisations: a straight high tone throughout the vowel or a linear fall. Listeners were asked to categorise the first vowel of the words as either "short" or "long". The type of tone did not affect listeners' categorisation when this vowel had extreme durations, but it had a significant effect at the three intermediate durations: the falling tone acted as a strong cue towards perceiving the first syllable vowel as "long". Vainio, Aalto, Järvikivi & Suni (2006) had previously observed that in CVCV words there was a static high tone on the first syllable whereas CVVCV and CVCCV words had a dynamic falling tone; the terms "static tone" and "falling tone" refer

to Xu's (2005) Target Approximation framework. This seems to be at variance with our results of tonal uniformity on the basis of which one would expect a rise during the first syllable of CVCV words and a rise-fall on that of CVVCV and CVCCV words. At the time of writing this it is unclear what causes the difference; it is possible that we are dealing with a difference between two varieties of SSF.

9.4 Segment durations and a speech timing model

In this section we review our observations of segment durations in the light of a speech timing model; the review is based on Suomi (submitted) in which the relevant statistical tests were reported. Our findings on segment durations in Finnish are in good agreement with the domain-and-locus model of speech timing proposed by White (2002), which is based on durational investigations of English. We shall first briefly describe the model, and then relate our findings to this framework. In White's model, domain refers to the prosodic constituent within which a timing process is operative, and locus refers to the particular segments that are affected by the process. The model comprises domain-head and domain-edge lengthening processes (the references to English durational patterns below are White's observations). Accentual lengthening is an example of a domain-head lengthening process, and domain-edge processes lengthen segments near the initial and final boundaries of constituents, e.g. phrase-final lengthening; the phrase is here the domain, and the locus begins with the final stressed syllable and continues to the phrase boundary (in English at least). The framework also recognises domain-span shortening (or compression) processes. To the extent that such processes exist, they are due to an inverse relationship between the size of some constituent and the duration of some subconstituent: for example, if the duration of a syllable decreases as word length (number of constituent syllables) increases, there is word-span compression.

The model assumes that each process is associated with a locus defined in phonological terms, and that processes may be distinguished by their distinct loci. According to the model, speech timing consists of localised effects: segments are produced with durations simply determined by intrinsic factors, modulated according to speech rate, until a locus of some timing process is reached. At this point, some extra amount of duration is allocated to the locus, with no regard paid to the segmental composition of the locus. This lengthening is distributed within the locus according to the structure and the segmental composition of the locus.

As a result of both of these factors, the degree of lengthening will not be evenly distributed within a locus. Thus e.g. in English, although the word appears to be the locus of accentual lengthening, there is more lengthening at the edges of the word than in its centre, and more elastic consonants may be lengthened more than less elastic segments. White concludes, for English, that “there is *no* unit into which an utterance may be exhaustively parsed that consistently imposes timing constraints upon its subconstituents. This is in contradiction to many theoretical accounts of speech timing which propose that there are units which mediate between linguistic structure and segmental duration” (p. 285, emphasis in the original). As far as we can judge, this conclusion holds for Finnish, too. In the following, findings of Finnish speech timing consistent with the model are discussed, without necessarily pointing out how these findings contradict theories that postulate such mediating units. Findings supporting the latter theories would also be discussed, but there do not seem to be any such findings.

To start with accentual lengthening, a domain-head lengthening process, White’s model thus suggests that the accented word receives a fixed amount of additional duration, and that the added duration is spread out within the locus of lengthening. Our results on accentual lengthening in Finnish are in complete agreement with this suggestion. Suomi (submitted) statistically re-analysed the data of earlier studies from this previously overlooked perspective. These analyses revealed that in Suomi (2005), the amount of accentual lengthening was statistically the same across the four word structures exemplified by *kana, kanta, kate, katse*. Similarly, in Suomi (2007), the amount of accentual lengthening was statistically the same across the monosyllabic to tetrasyllabic words, in which one set of words had a monomoraic first syllable (e.g. *se, setä, Setälä, Setälästä),* the other set a dimoraic first syllable (e.g. *sei, Seiko, Seikola, Seikolasta)*. Accentual lengthening was observed to extend from word onset to the end of the third syllable, with minor (statistically significant) lengthening appearing on the first segment of the fourth syllable. Again, as predicted by White’s model, all word types received the same amount of accentual lengthening, irrespective of the number of constituent syllables. The as yet unpublished results of the third author with similar materials indicate that the same situation (statistically equal amounts of additional duration irrespective of word length) also holds in three local varieties of SSF (one of which is Northern Finnish). Finally, in Suomi *et al*. (2003), the total amount of accentual lengthening was 91 ms in the CVCV words, 83 ms in the CVVCV words, and 99 ms in the CVCVV words. Here the figures,

calculated from Table 1 in the paper, are rather similar to each other and thus seemingly consistent with the prediction but, unfortunately, no statistical tests are possible as the original data have been lost.

Recall from above that the distribution of accentual lengthening in Suomi (2007) was highly non-linear in that C_1, M_1 and M_2 were extensively lengthened (between 75% and 58%), other segments less (19%), and that, proportionally, this pattern of distribution was a replication of that observed by Suomi (2005). Thus, as suggested by White's model, the degree of lengthening was not evenly distributed within the locus of accentual lengthening but was, instead, distributed according to the structure of the locus, and it is most probably a language-specific structural peculiarity of Finnish that precisely M_1 and M_2 are extensively lengthened (in addition to C_1), whether or not a non-moraic segment intervenes.

Suomi (submitted) argued that if the overall motivation for the durational alternations (for e.g. single vowels, the alternations between the four duration degrees [very short], [short] etc.) is to guarantee the tonal and temporal uniformity of the accentual rise-fall tune, then one would expect that especially those stretches of words tend to have equal durations during which the accentual tune is mainly realised. In the word structures examined (*pata*, *patsa*, *patna* etc., see above), this stretch consists of the sequence VCV (or M_1CM_2) in the C<u>VCV</u> words and of the sequence VCC (M_1M_2C) in the C<u>VCC</u>V words, i.e. both sequences consist of M_1 and the next two segments one of which is M_2. Previous research had shown that the accentual fall has reached very nearly the same phase by the end of VCV and VCC sequences. When M_1 and M_2 are contiguous in the first syllable, the F_0 fall has not yet reached as low a value by the end of M_2 as it has when M_2 is in the second syllable preceded by a non-moraic consonant. In the contiguous (VCC) case, the F_0 fall reaches the comparable low point during the consonant following M_2. The exact boundaries of the relevant tonal stretches may not coincide with segment boundaries, but the segmental stretches VCV and VCC seem to be a close approximation of the stretch during which F_0 behaviour is identical irrespective of structure.

The total mean duration of the VCV sequence in all C<u>VCV</u> words was 339 ms, that of the VCC sequence in all C<u>VCC</u>V words 328 ms. This difference was statistically significant, but the 11 ms difference was much less than the difference in the total durations of the CVCV and CVCCV words (57 ms). In the test words, durational differences among consonants with different intrinsic durations were compensated by the durations of other segments, and the finding

that the difference between the VCV and VCC sequences was only 11 ms strongly suggests that the compensatory effects were not so much aiming at equal total word durations, but rather to equal durations of the VCV and VCC sequences. Thus although the compensatory elastic behaviour did not result in perfectly identical durations of the two sequences, it greatly reduced the difference in these durations, relative to the differences in the total word durations.

Suomi (submitted) defined the VCV and VCC sequences as specific instances of a locus à la White (2002), arguing that the locus is a sequence of segments within which segment durations are adjusted in such a way that the total duration of the locus will be (approximately) the same irrespective of its segmental composition, that the purpose of the durational adjustments is to guarantee the uniform realisation of the accentual tune, and consequently named this locus the locus of duration-to-tone adjustments.

Within the locus of accentual lengthening, then, there is the shorter locus of duration-to-tone adjustments. Its domain is the word, and it consists of M_1 and the next two segments one of which is M_2. As has been shown above, vowels have systematically longer durations within this locus (single vowels are either [short], [longish] or [long], depending on the moraic structure) than outside it in the same foot (where they are always [very short]), so it can be said that a constant amount of extra duration has been allocated to the locus, relative to the duration of corresponding segment sequences outside the locus. And the duration is distributed differently according to the segmental/moraic structure of the locus. For example, in Suomi (submitted), if C_2 in both CVCV and CVCCV word was /t̪/, an intrinsically long consonant, then V_1 had a shorter duration than when C_2 was /l/ or /m/, intrinsically shorter consonants; this was one of the elastic compensatory effects observed.

Suomi (submitted) further argued that, given its functional motivation, the locus of duration-to-tone adjustments should be observable in all word structures in which it can in principle be observed, and statistically re-analysed data from previous experiments from this perspective. This excludes short word structures like CVV, in which the definition of the locus is not fulfilled; Suomi (2007) observed that in precisely the CVV words, a large part of the accentual fall continued in the first syllable of the next word, in contrast to longer words fulfilling the definition of the locus (and the CV function words behaved very much differently from the at least dimoraic words). Suomi (submitted) established

that, in Suomi & Ylitalo (2004), the loci in CVCVCV, CVCVCCV, CVCCVCV, CVCCVCCV, CVVCVCVV word structures were statistically non-distinct.

In the CVVCVVCVV word structure the locus had a longer duration than it did in those structures just mentioned. But the deviant behaviour of this structure has an independent explanation: the last segment of the locus is an onset consonant followed by a double vowel, a context in which a consonant is known to have a systematically longer duration than it does when followed by a single vowel, as discussed above. Apart from this independently motivated counterexample, then, the loci consisting of VCV, VCC and VVC had non-distinct durations. There is thus evidence that the concept of a locus of duration-to-tone adjustments, a locus with a nearly constant duration irrespective of word structure, can be generalised to all word structures so far investigated. Notice however that Suomi & Ylitalo (2004) studied unaccented words. Here again, then, unaccented words are durationally prepared to carry accent: recall from section 9.3 above Laurence White's conclusion that, in Finnish, preliminary durational adjustments are necessary even in the absence of accent to allow the accentual uniformity.

Both of the domain-head lengthening processes just discussed, accentual lengthening and the process of duration-to-tone adjustments, have the word as the domain. But they have distinct loci. The locus of accentual lengthening consists, somewhat tentatively, of the word's first three syllables (where applicable), including the word-initial consonant. The shorter locus of duration-to-tone adjustments consists of M_1 and the next two segments one of which is M_2. The two processes have functional unity in that both contribute to the realisation of accentual prominence: accentual lengthening signals contrastive accent durationally, and the duration-to-tone adjustments enable the uniformity of the accentual tune across different word structures. Within both loci, the extra amount of duration is distributed differently according to the structure and the segmental composition of the locus, consistent with White's model.

Domain-span shortening processes are due to an inverse relationship between the size of some constituent and the duration of some subconstituent. White's model excludes such processes in English, and there are no findings known to us that suggest the existence of such processes in Finnish. To be sure, Iivonen (1974) did observe shortening of segment durations as word length was increased, but he studied single-word utterances of increasing length, and thus utterance length co-varied with word length. Consequently, this finding is ambiguous as to whether

the effect is at the word level or some higher level, and as to what type of process is involved. If the words were contrastively accented (this is not specified), then some durational adjustments relating to accentual lengthening may have been operative. And, perhaps most importantly, it is not possible in single-word utterances to distinguish initial and/or final lengthening from some sort of domain-span process. Thus, this single-word study is open to multiple interpretations. Consequently there is, to our knowledge, no unambiguous evidence of domain-span shortening processes in Finnish, but there are many counterexamples.

Let us first consider potential domain-span compression at the syllable level. In Suomi (submitted), it was observed that a single vowel had a longer (namely [longish]) duration in the first syllable of CVCCV words than in that of CVCV words (where it was [short]), a finding consistent with previous ones. It was also observed that /l/, /m/ and /t̪/ all had a longer duration acting as C_2 in the CVC_2CV words than in the CVC_2V words, a finding similarly consistent with previous findings. In both of these phenomena, a segment has a longer duration in a longer syllable (i.e. one with more segmental material) than in a shorter one, contrary to domain-span shortening. Notice, in passing, that CVCV words and CVCCV words both also constitute a foot, and that there are good grounds for arguing that, as a result of the alternations, the total durations of these words are more similar than they would be without any alternations. But these word structures are a special case, they are minimal disyllabic words capable of containing the duration-to-tone adjustments locus, and the fact that they also constitute a foot is a mere coincidence that does not prove that there is a general tendency towards foot isochrony. Apart from this special case with an independent explanation, feet do not exhibit such a tendency (see the next two paragraphs).

To take another counterexample to putative domain-span shortening processes, recall from above that Suomi & Ylitalo (2004: 58) observed that each of the consonants in CVV.CVV.CVV words had reliably and on average 12 ms longer durations than those in CV.CV.CV words (both sets of words also constitute feet), and that the authors computed that in Lehtonen (1970) the corresponding reliable difference in five structural pairs compared was on average 13 ms. In this phenomenon, longer duration of C in CVV than in CV syllables, segments again have longer durations in syllables with more phoneme segments than in those with fewer ones. As a consequence, the durations of feet and words are also boosted accordingly. Instead of domain-span shortening, then, there is

bottom-up accumulation of duration to higher-level units, from segments to syllables to feet/words.

O'Dell & Nieminen (2006) studied foot timing in Finnish in the framework of their oscillator model. The model consists of a dynamic system of many mutually synchronising hierarchical oscillators (e.g. feet consisting of syllables), and the model predicts domain-span shortening as the default situation. Three of the five speakers in the experiment exhibited no signs of foot timing, and the results for the other two speakers were partly consistent and partly inconsistent with predictions of foot timing (or data had to be discarded because of non-fluent pronunciation).

Suomi (2007) found little support for polysyllabic shortening (Lehiste, 1972); that is, domain-span compression at the word level. Lehiste suggested that, as the number of constituent syllables of a word increases, mean syllable duration decreases. In Suomi (2007), in both unaccented and contrastively accented words spoken in utterances of approximately equal length, increase in word length did not systematically shorten the duration of constituent syllables, a finding consistent with the earlier one by Lehtonen (1974) who studied Central Finnish. Thus in the statistical analyses (with Bonferroni correction), three comparisons were consistent with polysyllabic shortening, 31 failed to exhibit a reliable difference, and two were contrary to polysyllabic shortening. Thus, there was little evidence for polysyllabic shortening *(alias* a tendency towards word isochrony). The as yet unpublished results by the third author of this book confirm this for Northern Finnish, and show that polysyllabic shortening is not operative in the other two varieties, either. Thus, altogether, the absence of polysyllabic shortening has been demonstrated in four local varieties of SSF.

To be sure, in the contrastively accented words in Figures 1 and 2 in Suomi (2007), there were numerical trends observable such that syllable durations were usually slightly shorter in longer than in shorter words. These trends are not likely to be random. Recall from above that all contrastively accented words in Suomi (2007) (and in the other relevant studies reviewed) received statistically the same amount of accentual lengthening irrespective of word length. This means, in principle, that the longer the word, the less of accentual lengthening was received by given constituent syllable. This trend is due to the way accentual lengthening works (constant amount of lengthening received by accented word irrespective of number of constituent syllables, less of the lengthening allotted to a given syllable in longer words). This effect is called polysyllabic accent effect by White (2002).

The vocalic portion of *Sei* words had a statistically longer duration than the same portion had in *Seikola* and *Seikolasta* words (when Bonferroni correction was used). Thus there was some evidence of the polysyllabic accent effect. But it is important to maintain a clear distinction between polysyllabic shortening and the polysyllabic accent effect: observation of the latter is no evidence for the former.

The borderline between word-level and utterance level prosody may be like a line drawn in the water. Thus e.g. foot timing, if it exists, may cross word boundaries, at least in languages that do not have fixed initial stress. The durational effect to be discussed next, utterance-final lengthening, clearly belongs to utterance-level prosody, but it is nevertheless discussed here, in the context of other durational effects. In White's model, phrase-final lengthening is a domain-edge process whose domain is the phrase. We assume that 'phrase' may in this context denote any constituent larger than the word, and that utterance-final lengthening is an instance of phrase-final lengthening. To our knowledge, final lengthening in 'phrases' shorter than the utterance (e.g. in phrases proper) has not been systematically investigated in Finnish.

What the locus of utterance-final lengthening is in Finnish, is unclear as this phenomenon has been only little investigated. Myers & Hansen (2006) observed that both single and double vowels were extensively lengthened in utterance-final position when [h]-like voiceless endings of modally phonated vocalic portions, endings typical of utterance-final vowels in Finnish, were included in vowel durations: single vowels had 66% longer and double vowels 52% longer duration than in utterance-medial position. But when only the voiced portions of the utterance-final vowels were considered, single vowels were not lengthened at all, double vowels 20%. Myers & Hansen's further perception experiments showed that the voiceless vowel ending, which accounted for most of the utterance-final lengthening of single vowels, does not contribute to Finnish listeners' perception of vowel quantity. This suggests that Finnish avoids durational overlap between the voiced portions of single and double vowels.

However, estimates of the magnitude of utterance-final lengthening in Myers & Hansen's material may be conservative. For example, the final /ɑ/ in *koira* represented the final position, and the same vowel in *koirako* (*-ko* is a question marker) represented the non-final position. It is possible that the /ɑ/ in the latter word is within the locus of utterance-final lengthening. If it is, then it must have undergone some degree of utterance-final lengthening, which leads to a smaller

estimate of lengthening of the final vowel than would a comparison with segments further away from the utterance boundary.

Nakai, Kunnari, Turk, Suomi & Ylitalo (to appear) studied utterance-final lengthening in Finnish CV, CVV, CVCV, CVCVV, CVVCV and CVVCVV words. In the following, we look at the results obtained using the authors' segmentation criterion 'V(V) voicing', according to which the voiceless end of the utterance-final vowels was not included in vowel durations; any periods of breathy phonation were included. The authors observed utterance-final lengthening to extend, backwards from utterance offset, to the initial syllable of the disyllabic words, which supports the assumption that Myers & Hansen's estimates of utterance-final lengthening were conservative. Altogether Nakai *et al*. observed that the first-syllable double vowels were reliably lengthened but first-syllable single vowels were not, word-medial consonants (onsets of the second syllable) and word-final double vowels were lengthened, as were word-final single vowels in CVVC$\underline{\text{V}}$ words. But word-final single vowels in CVC$\underline{\text{V}}$ words were not lengthened. The precise locus of utterance-final lengthening in Finnish, as well as the question of how the lengthening is distributed within the locus, need further clarification, but as concerns the results vis-à-vis the quantity opposition, they can be summarised as obeying the principle: lengthen segment durations only if the lengthening does not jeopardise the quantity opposition. Of course, double vowels can be lengthened *ad libitum* without any such risk, the medial single consonants can similarly be lengthened considerably without such a risk, as can the [very short] second-syllable single vowel in CVVC$\underline{\text{V}}$ words (this vowel being [very short]). These are all segments that were considerably lengthened in utterance-final words. But the [long] word-final single vowel in CVC$\underline{\text{V}}$ was not, because if it were, there would not be a safe margin against the CVCVV word structure; recall from above (section 9.3) that the CVCV and CVCVV word structures are not optimally distinguished, even without utterance-final lengthening. Thus even though there is utterance-final lengthening in Finnish, it is constrained by demands not to neutralise the quantity opposition.

Altogether, then, there are strong empirical grounds for arguing that, outside specific loci, speech timing in Finnish is determined bottom-up. There do not seem to be units that mediate between linguistic structure and segmental duration at all points in the speech string. Outside loci, the syllable does not impose durational constraints on its constituent segments, the foot does not determine the durations of its constituent syllables, etc. Instead, outside specific loci, the

durations of phonological constituents increase as a linear function of the durations of their subconstituents.

To us, White's (2002) domain-and-locus model has been invaluable in inspiring us to ask questions that we would otherwise not have thought of. It is remarkable that a timing model based on English is so consistent with observations in a full-fledged quantity language like Finnish. Such a language is a rather stringent testing ground for speech timing models.

10 Utterance-level prosody

As already mentioned above, there is no clear distinction between word-level and utterance-level prosody. For example, single-word utterances, sometimes assumed to exhibit the canonical lexical properties of the words occurring in such utterances and seemingly lacking utterance-level effects, nevertheless are utterances, with consequent prosodic effects. Single-word utterances are usually focused, they carry the sentence accent, they are subject to domain-edge durational effects, etc. In the preceding Chapter, some utterance-level prosodic durational effects were discussed. In this Chapter, more unquestionably utterance-level prosodic phenomena are discussed.

10.1 Orientation

"*There is very little intonation in spoken Finnish…*" This is a stereotype that is both age-old and ever-new, although its world-weary and hip intellectual reification tends to be somewhat mock-serious and meta-conscious nowadays (see e.g. the views of Mr. Leevi Lehto, a well-known Finnish poet and translator of English literature: http://www.leevilehto.net). There have indeed been some peculiar, if not quite hilarious, misunderstandings about the nature and functions of prosody in spoken Finnish, particularly at the utterance-level. It has been claimed, for instance, that there is no observable intonation in the speech of Finnish males (Brazil, Coulthard and Johns 1980). On the other hand, as for accentuation, it was once believed that in spoken Finnish the first item of an intonation-group (tone-group) is accented — presumably regardless of the semantic structure of the utterance (Heringer & Wolontis 1972). One may wonder to what extent such pre-conceived ideas reflect the national stereotype of the "silent Finn" (incidentally, still perpetuated in Kaurismäki's films world-wide). However, in more objective descriptions of all things Finnish, it is now widely recognised that intonation (e.g. rising intonation) serves quite specific functions in Finnish though it may be true that these distinctions are used less systematically in Finnish than in some other languages (Välimaa-Blum 1993).

In the following sections, we shall look at accentuation patterns in Finnish, as well as the semantic/pragmatic role of intonation, the issue of rising intonation being particularly interesting from the communicative viewpoint. We will also

look in some detail into the interrelationship between prosody (intonation and voice quality) and emotions and attitudes in Finnish.

10.2 Accent and information structure

In spoken Finnish, intonational phenomena happen in chunks of speech that have been labelled as *rytmijakso* ('rhythm group'), *lausuma* or *puheke* ('utterance'), and *fonologinen fraasi* ('phonological phrase'). Such a group typically consists of a few words (rarely more than seven or eight in non-prepared speech), and the unit is often (but not always) preceded and followed by a pause. Within such a group, the main accent typically, or by default, falls on the last (lexical) item. This accent placement generally presupposes that the whole intonation-group containing all-new information is in focus, i.e. it gets a rhematic accent. The accent falling on the last item thus reflects a rhematic focus.

In the YLE (Finnish Broadcast Company) news on 6 February 2008 the news anchor announced: *Pääministeri Vanhanen on matkustanut Intiaan* ('Prime Minister Vanhanen has travelled to India'). This piece of news was all-new, as it were, containing a rhematic focus and thus a rhematic accent on *Intiaan*. Perceptually, one could hear a slight accent on *Intiaan*, carrying a rheme-signalling focus (in effect, the intonation pattern on *Intiaan* was a gentle rise-fall, as with the other lexical items in the utterance carrying thematic accents, i.e. even less prominent rise-falls).

If the news anchor had mistakenly announced that Mr. Vanhanen had travelled to Pakistan, he might have corrected this by saying: *Anteeksi*, *pääministeri Vanhanen on matkustanut Intiaan* ('Excuse me, Prime Minister Vanhanen has travelled to India') with a more pronounced (rising-falling) intonation on *Intiaan*, signalling a contrastive accent. A contrastive accent of this kind would usually be accompanied also by increased duration.

Finally, an emphatic accent might be placed on *Intiaan* if someone (e.g. a Finnish anti-Hindu activist) was expressing strong indignation about the destination of the journey; in such a case, F_0 features, intensity and duration would all be further boosted (and possibly each syllable in *Intiaan* would get a separate emphatic rising-falling tone).

Now, it can be argued that there are at least three (and possibly four) types of accent in spoken Finnish. As was shown earlier in this book, mere word stress is not realised tonally. Accent, on the other hand, is realised at least tonally. A

thematic accent is realised as a gentle rise-fall, typically falling on lexical items (content words) in an intonation group (*pääministeri*, *Vanhanen* and *matkustanut* in *Pääministeri Vanhanen on matkustanut Intiaan* — note that *on*, here meaning 'has', carries mere word stress here). A rhematic accent is then realised as a more prominent rise-fall (*Intiaan* in *Pääministeri Vanhanen on matkustanut Intiaan*). These two degrees of accent are not realised durationally. A contrastive accent, on the other hand, is realised as an even more prominent rise-fall with increased segmental duration (on *Intiaan* in *Anteeksi*, *pääministeri Vanhanen on matkustanut Intiaan*). The fourth degree of accent, the emphatic accent, is not a phonological phenomenon as such as it reflects the degree of emotion rather than the degree of contrast in a speech situation. With the emphatic accent, all prosodic features (F_0, intensity, duration) can increase "unlimitedly" (relatively speaking) in unison with the speaker's affective state (reflecting indignation, enthusiasm, surprise, etc.).

The examples above with different degrees of accentuation can be displayed typographically as follows:

Pääministeri Vanhanen *on* ***matkustanut INTIAAN.***

Anteeksi, pääministeri Vanhanen *on* ***matkustanut INTIAAN.***

En *kyllä* ***usko*** *että* ***pääministeri Vanhanen*** *on* ***matkustanut IN TI AAN.***

('I really do not believe that Prime Minister Vanhanen has travelled to India')

In these examples, the words in regular font, generally representing grammatical/functional items, are unaccented, merely bold-faced words are thematically accented, capitalised and bold-faced words are rhematically accented, capitalised and bold-faced words in a bigger font are contrastively accented, and words in the biggest font are emphatically accented.

A rhematic accent reflects a rhematic focus, i.e. a situation in which "everything is new", in the sense that clearly old information (earlier mentioned information) is not available. Often, of course, a number of items in speech have been mentioned earlier, or can be interpreted as old information on the strength of context, inference, etc. In the YLE newscast cited above, the news anchor announced, at a later point: *Vanhanen on useaan otteeseen kommentoinut Kenian-kriisiä Intian-matkansa aikana* ('Mr. Vanhanen has several times during his visit

to India commented on the crisis in Kenya'). Orthographically, the utterance and its accentuation pattern can be represented as follows:

> ***Vanhanen*** *on* ***useaan*** *otteeseen* ***kommentoinut KENIAN-KRIISIÄ Intian-matkansa*** *aikana****.***

In this utterance, the main accent was observable on *Kenian-kriisiä* ('crisis in Kenya'); there was thus a rising-falling accent on *Kenian-kriisiä* (this rhematic-accent was realised tonally). In the context, *Intian-matka* (visit in India) was undoubtedly old (given) information, and it only received a thematic accent. This utterance therefore represents a thematic focus (note that even such an utterance contains a rhematic accent in that extant part of the utterance which conveys new information). In Finnish, accent placement is governed by the information structure of the utterance; if the utterance contains old information toward the end, the main accent is assigned to an item in an earlier position containing rhematic information. The principles of accent assignment and rhematic vs. thematic focus are similar to those found in most languages (e.g. English, German). Although the distribution of information does govern accent placement at some general level (in Finnish and in languages universally), we should not forget that the information structure of an utterance is ultimately a cognitive process, known only to the speaker. There is thus some truth to Bolinger's (1972) thesis that "accent is predictable if you are a mind reader".

Bolinger's thesis brings to mind another spreading mannerism, the accentuation of function words, or other words, when the information structure of the utterance obviously does not motivate the accentuation. For example, a voice on the radio recently announced that *Soitto**AIKA ON** päättynyt* 'the call**TIME** (time for making telephone calls to the program) **HAS** ended', with prominent accents on the capitalised words. In *soitto**AIKA*** accent is on the second part of a compound, ***ON*** is a finite verb form and, apart from mannerisms like these, finite verbs are never accented, except when they are contrastively accented. A normal, less annoying rendering of the announcement would be: ***Soitto**aika on **PÄÄTTYNYT***.

10.3 Intonation

Neutrally uttered complete statements in Finnish generally take a smoothly descending pitch contour; the first syllable is uttered somewhere above (or at) the

middle of the speaker's voice range, and the last syllable is uttered on a very low pitch (often, the end of the intonation-group is accompanied by creak). This clearly seems to be the commonest pitch pattern in non-emotional Finnish (Iivonen 1998), and it has been documented over a long period of time (Peltonen 1901; Sovijärvi 1956; Hirvonen 1970). Incidentally, many authors have also generally claimed in this context that Finnish is "flat" and "monotonous" (Peltonen 1901; Sovijärvi 1956; Monola 1976). Finnish intonation, at least the most "neutral pattern", can be described as a succession of declining rising-falling patterns (on content words), with an end reaching a very low F_0 level (eventually containing non-modal phonation). The rhematic accent is an observable intonation phenomenon but does not stand out particularly markedly from the concatenation of declining rising-falling patterns. Contrastive/emphatic accents, on the other hand, can represent highly conspicuous aberrations (upward F_0 twists) from the succession of rising-falling F_0 patterns steadily reaching lower positions toward the end of the utterance. In a rhematic focus, the rhematic accent is usually near the end of the utterance as it is more "economical" to begin the utterance with old information and present the new information after the "introductory" phase. Since the rhematic accent is not normally highly prominent phonetically, the "standard" pattern in Finnish intonation is indeed a steadily and smoothly declining succession of rising-falling F_0 patterns — hence, possibly, the stereotype about the paucity of intonation contours in Finnish.

It has been claimed (Hirvonen 1970), with empirical evidence backing up the argument, that in spoken Finnish "communication proper", i.e. statements, and "communication with an appeal to the listener", i.e. questions and commands, differ from each other intonationally in terms of the height of the initial pitch level: statements begin around the middle of the speaker's pitch tessitura (as was discussed above) while questions and commands have a higher initial pitch. Indeed, if a tape-recorded Finnish question is played backwards, one is likely to hear a rising intonation! This feature of Finnish intonation possibly reflects a prosodic language universal: it has been reported that, across languages, questions typically have a higher overall pitch than statements, and, importantly, the higher pitch need not occur near the end of the intonation-group (Bolinger 1989). Interestingly, for English, a similar phenomenon was observed very early (Hart 1551, 1569): "… their [the interrogative and admirative] tunes doe differ from our other maner of pronunciation [statements] at the beginning of the sentence."

For spoken Finnish, an explanation of the higher initial pitch has to do with the distribution of accents in the utterance. Let us consider three utterances: *Tulen teille illalla* ('I'm coming to your place tonight'), *Tule meille illalla*! ('Come to our place tonight!') and *Tuletko meille illalla*? ('Are you coming to our place tonight?'). In the statement, the initial pitch level is in all likelihood lower than in the command and in the question. In the statement, *tulen* and *teille* get a thematic accent, while *illalla* gets a rhematic accent. As for the command (or exhortation), it can be argued that the first item, *tule*, represents rhematic information, and thus gets a rhematic accent. In the question, on the other hand, the item to which the interrogative particle -*ko* suffixed, i.e. *tule*-, represents rhematic information, and therefore also gets a rhematic accent. The pitch height difference between statement and command/question can now be explained with reference to accentuation: in the command and the question, the rhematic accent is in an earlier position than in the statement. Schematically, the situation can be described as follows:

Tulen teille ILLALLA. ('I'm coming to your place tonight').

TULE meille illalla*!* ('Come to our place tonight!').

TULETKO meille illalla? ('Are you coming to our place tonight?').

It should be pointed out here, as before, that, in spontaneous freely floating conversation and "chit-chat", it is not possible to predict the accent pattern in any deterministic manner. Nevertheless, schematic rules presented here can be a first approximation of the situation (even if conversation analysts may take issue).

Finally, it can be pointed out that so-called progredient intonation (continuation intonation) is quite common: the pitch level at the end of the utterance remains around the middle of the F_0 range, without a fall. In fact, the end may contain a slight rise. Intonation of this type typically occurs in a speech situation where something is left open or inconclusive, for example:

Sovitaanko asia näin vai… ('Can we agree on this or…')

In this example, (*näin*) *vai* would probably remain at a steady (or rising) pitch level clearly above the low part of the speaker's voice range.

10.4 Rising intonation

Whether or not there are rising intonation contours is Finnish has been a matter of storm in the proverbial teacup. Traditionally, rising intonation has been described as alien to Finnish intonation, or even downright unacceptable (Peltonen 1901; Marjanen 1932; Ahonen-Mäkelä 1975). However, it seems that while rising intonation is clearly less common than falling intonation, no-one can really claim that rising intonation *never* occurs in Finnish. Any speaker of Finnish, a native or second-language speaker, has certainly observed speech situations involving clearly rising intonation in (colloquial) Finnish.

A clear case of rising intonation in Finnish is what is known as an echo-question, basically repeating the previous utterance (and possibly conveying disbelief). An example would be:

Hän siis sai synttärilahjaksi mitä? ('He got a what as a birthday present?')

Mitä? ('What?')

In these examples, *mitä* might carry a steeply and rapidly rising F_0 pattern.

Another case is small function words serving interactional and politeness purposes. A rising intonation may occur on one-word expressions securing a minimum amount of common ground between the speaker and hearer:

HuomenTA! ('Good morning!')

PäiVÄÄ! ('Good afternoon!')

KiiTOS! ('Thank you!')

With such expressions, the accent is assigned to the last syllable, with a rising intonation. A sense of socially required bond or rapport is created with these lexico-prosodic patterns. A different case would be *KIItos*!, with an accent on the first syllable and a falling tone: an overtone of genuine indebtedness would be created.

An analogous case would be *Anteeksi* uttered with different accentuation and intonation patterns. With an accent on the last syllable and a rise, the effect is 'Sorry, I could not hear you':

AnTEEKS!

In contrast, with an accent on the first syllable and a fall, the paraphrase might be something like: 'I am so truly sorry':

ANteeksi!

The first type of expression was, incidentally, used in a riotously funny TV sketch where a hideously incompetent construction worker repeatedly, but apparently quite accidentally, bulldozed down families' houses and commented afterwards: *Hö*, *AnTEEKS*! (freely translated as 'Oops, sorry, no worries!').

The examples above represent lexico-prosodic structures, as it were, since the fusion between intonation and lexical structure is more or less fixed (at least for those speakers who regularly use these patterns). In modern spoken Finnish, at least in the parlance of adolescent females in the Helsinki area, (high) rising intonation is becoming more prevalent also in expressions showing more creative language use. It seems that rising intonation is becoming more common in contexts where the speaker wishes to indicate that he/she assumes that the topic represents a shared world between the conversationists and that much can be taken for granted, and much is already mutually understood, in the speech situation. The following extract was heard in an enthusiastic and rambling conversation between two high school girls:

> *...se* [Michael Emerson] *on siis Lostissa se tutkija siis siellä toisella saarella niiden toisten pomo*
>
> …'he [Michael Emerson] on Lost is the researcher on the other island you know the boss of the others'

The intonation on *tutkija* 'researcher' and *pomo* 'boss' was unequivocally a (high) rise, perceptually very similar to (American) English question intonation. The speakers were probably familiar with the topic (the massively popular *Lost* TV show) and the characters and actors. Thus the current speaker could, perhaps, assume that the epithets associated with Michael Emerson, the actor, were known to the interlocutor. By using a rising tone on these items, the speaker then wished to indicate that she was not really saying anything that the other speaker did not already know. Alternatively, it is possible that the speaker wished to indicate that

the other speaker could judge whether the information she was offering was relevant and/or useful. Thus the speaker wished to give the interlocutor leeway as to how to react to the presented information. The interactional role of the (high) rise could be glossed as: 'I am giving you these pieces of news, please feel free to tell me what you think'.

Another authentic example was heard in a context where two students (females in their twenties) discussed the possibilities of a TV dinner:

Onks mitään kunnon ruokaa?

No on… pasteijoita.

'Is there any OK food?'

'Well there are some… pastries.'

The intonation on *pasteijoita* ('pastries') was clearly a rising tone. Again, it seems that the speaker wanted to solicit the interlocutor's judgement about the quality of pastries as food in the context. An attitudinal interpretation might be: 'This is what I have got, tell me if this is OK'.

By using the (high) rise in contexts like these, the speaker probably also wishes to offer a face-saving discoursal resource: a rise takes off the edge of a statement and the other speaker is given an opportunity to view the discourse horizon before reacting. If necessary, both speakers can bale out of the topic/statement/proposition.

According to Routarinne (2003), rising intonation is becoming more and more common in the speech of young women particularly in the Finnish capital area (Helsinki and its surroundings). One may speculate that (young) female speakers are more sensitive to the issue of face-saving in conversation, thus preferring (high) rising intonation on items potentially causing uncertainty or even disagreement. To use rising intonation in such contexts may then be a way to pre-empt such potential clashes. It can also be speculated that a (high) rising tone in Finnish declarative utterances can be seen as reflecting two conditions: the scalar relationship between an item and the context, and the "uncertainty" of the speaker with respect to the relevance of the item on the particular scale (cf. *pastries*/*OK food* in the example above).

Recall from Chapter 2 above that creaky phonation characterising speech irrespective of prosodic position is an increasingly common property of young,

especially female speakers, and possibly in the Finnish capital area in particular. We have no evidence that the two changes in speaking style, rising intonation and overall creaky phonation, go hand in hand, but this is possible. Nevertheless, in both trends, young female speakers seem to be the forerunners of change. This is not counter to how manners of speaking (including sound changes) or other human fashions often originate and spread.

It would be interesting to conjecture about the origins of the (high) rising intonation in spoken Finnish. Is it indigenous or a result of external influence? It has been well documented (Cruttenden 1997) that high rising terminals are common in Pacific Rim English (spoken on the west coast in the USA and Canada and on the eastern coast in Australia). High rise terminals generally convey different kinds of nuances of uncertainty and openness in declarative utterances in PR English, to the extent that it has been lamented that some speakers of English do not seem to be sure on anything any more, even of their own names. If someone introduces himself as "*John Smith*?" in LA, offering reassurance is not required.

Finally, we should not forget that emotional speech is very likely call for (high) rising intonation contours in Finnish. Utterances of annoyance, disbelief, shock, etc. may carry steeply rising wide intonations (we will return to this below).

10.5 Descriptive frameworks for Finnish intonation

Traditionally, Finnish intonation has been described with common labels such as "fall", "rise", "rise-fall", etc., directly labelling the essential pitch pattern. This framework is still being used (Iivonen 1998). Basically, the descriptive system is similar to that used in the traditional "British school" of intonation studies (e.g. O'Connor & Arnold 1973). The intonation group is assumed to have a fairly clearly-defined internal structure, containing at least the tonic/nuclear syllable (i.e. the main accent, often a rhematic accent), with optional proclitic and enclitic elements. In the analysis of speech data, the following tones are generally thought to be possible: fall, rise-fall, rise, fall-rise, and level tone. Above, we have discussed most of these tones in the Finnish intonation context. In the traditional type of intonation analysis, the phrase final (tone unit final, intonation group final) tone choices are investigated in detail. The aim is usually to study the most salient pitch pattern at the end of each utterance.

In intonation descriptions, the focus is on *nuclear tones* (as defined in the British school framework of intonation analysis) or on *nuclear accents* occurring near the end of the intonation phrase (as defined in the ToBI framework). ToBI (Tones and Break Indices) is a framework for developing generally agreed-upon conventions for transcribing the intonation and prosodic structure of spoken utterances in a language variety (see e.g. Beckman & Ayers 1993). Originally developed for English, the system has been used in the description of the prosody of a number of languages, including Dutch, Spanish, Italian, Japanese and French. The ToBI model of intonation description is now something of a standard universally, and it is gaining ground in investigations of Finnish intonation. In this context, a brief summary of this system may be in order.

In the ToBI framework, pitch accents falling on the stressed syllables of semantically important words are marked with the star "*", and an intonation phrase may have several pitch accents (H*, L*, L*+H, L+H*, and !H). H* is a peak accent: the accented syllable is in the middle or upper part of the speaker's pitch range (this is the default accent, and there may be a slight subsequent fall). L* is a low accent where the accented syllable is in the lowest part of the speaker's pitch range (also a common accent type). L*+H is a scooped accent where the low tone on the accented syllable is immediately followed by a rise to the middle or upper part of the speaker's pitch range. L+H* is a rising peak accent where there is a high pitch on the target syllable after a steep rise. !H is a stepped accent where the accented syllable is in the middle or upper part of the speaker's pitch range, a step lower than the preceding H*. L- and H- are phrasal tones filling the interval between the last pitch accent and the final boundary tone (L- is the default, H- is semantically marked), while L% (default) and H% are final boundary tones. The boundary tones occur at each full intonation phrase boundary. Thus an intonational phrase, which may contain one or more intermediate phrases, ends with a boundary tone on its right edge. %H and %L (default) are initial boundary tones. Typically, full intonation phrases represent the following types: L-L% (default pattern in declaratives), L-H% (list pattern with non-final items), H-H% (yes-no question pattern), and H-L% (plateau).

There is no one-to-one relationship between the ToBI system and the nuclear tone description; some of the correspondences include: fall (H*L-L%), rise-fall (L*+H L-L%, L*+H H-L%), rise (L* H-H%), fall-rise (H* L-H%, H+L* H-H%), level tone (H* !H-L%).

To adapt ToBI to Finnish, we would like to suggest a modification; of course those who have developed ToBI are in no way responsible for this modification, and they have never claimed that ToBI could be applied to Finnish. Firstly, as discussed in section 9.2 above, the Finnish accentual tune, at least in Northern Finnish, is usually a rise-fall, with the fall clearly being an essential part of the tune. For this reason we have concluded that the tune is best formalised as the tritonal sequence LHL, rather than as a sequence of two tones. Secondly, both the initial L and the H are anchored to the initial, stressed syllable (to its onset and to M_1, respectively). Therefore, we see no need for starred tones to duplicate this information. Using this modified system, the most basic intonation pattern in Finnish (occurring on a neutral declarative utterance) can be described as follows; basically, a similar ”standard” Finnish utterance, with similar “standard” intonation, has been presented by Välimaa-Blum (1993). The basic contour is falling, with an accent on each content word. The pitch accent is, by default, LHL, there are two initial boundary tones, %L and %H, and two final boundary tones, L% and H%. A neutral declarative could contain the following pattern (“Laina lends Laina a loan”):

	Laina	*lainaa*	*Lainalle*	*lainan*.
%L	LHL	LHL	LHL	LHL L-L%

Analogously, a similar utterance with words containing one-moraic first syllables would seem to get the following type of pattern (“Late levels down the snow for Lulu”):

	Late	*lanaa*	*Lululle*	*lumet*.
%L	LHL	LHL	LHL	LHL L-L%

The nuclear accents (rhematic accents in these cases) would probably be on *lainan* and *lumet*. Notice that the two utterances are suggested to receive the same notation. This correctly captures the generalisation that the accentual tune is uniform irrespective of word structure. Only the rise is realised during the initial stressed syllable if this is monomoraic, and both the rise and a large part of the

fall are realised on the first syllable if this is at least dimoraic. This difference need not be explicitly shown, as it follows from the fact that the H tone is anchored to the end of M_1.

ToBI labelling is commonly used in the prosodic transcription of (British and American) English, and good inter-transcriber consistency can be achieved as long as the voice quality analysed represents normal (modal) phonation. Certain discourse situations, however, seem to consistently produce voice qualities different from modal phonation, and the prosodic analysis of such speech data with traditional ToBI labelling may be problematic. Typical examples are breathy, creaky and harsh voice qualities. Pitch analysis algorithms, which are used to produce a record of the fundamental frequency (F_0) contour of the utterance to aid the ToBI labelling, yield a messy or lacking F_0 track on non-modal voice segments. Non-modal voice qualities may represent habitual speaking styles or idiosyncrasies of speakers but they are often prosodic characteristics of emotional discourse (sadness, anger, etc.). As was mentioned earlier, spoken Finnish is often characterised by creak, especially in final position in declarative utterances; for many speakers, creak is quite wide-spread and consistent. Therefore, like some special (possibly emotion-specific) speech genres of English, spoken Finnish in general may be problematic for ToBI.

A potential modified system would be "4-Tone EVo", a ToBI-based framework for transcribing the prosody of modal/non-modal voice in (emotional) English (Toivanen 2006). As in the original ToBI system, intonation is transcribed as a sequence of pitch accents and boundary pitch movements (phrase accents and boundary tones). The original ToBI break index tier (with four strengths of boundaries) is also used. The fundamental difference between 4-Tone EVo and the original ToBI is that four main tones (H, L, h, l) are used instead of two (H, L). In 4-Tone EVo, H and L are high and low tones, respectively, as are "h" and "l", but "h" is a high tone with non-modal phonation and "l" a low tone with non-modal phonation. Basically, "h" is H without a clear pitch representation in the record of F_0 contour, and "l" is a similar variant of L.

The system has not so far been tested for Finnish but preliminary tests for (emotional) English have been made. To assess the usefulness of 4-Tone EVo, informal interviews with British exchange students (speakers of Southern British English) were used (with permission obtained from the subjects). The speakers described, among other things, their reactions to recent global tragedies (the emotional overtone was, predictably, rather low-keyed). The discussions were

recorded in a sound-treated room; the speakers' speech data were recorded directly to hard disk (44.1 kHz, 16 bit) using a high-quality microphone. The interaction was visually recorded with a high-quality digital video recorder directly facing the speaker. The speech data consisted of 574 orthographic words (82 utterances) produced by three female students (20–27 years old). Five Finnish students of linguistics/phonetics listened to the tapes and watched the video data; the subjects transcribed the data prosodically using 4-Tone EVo. The transcribers had been given a full training course in 4-Tone EVo style labelling. Each subject transcribed the material independently of one another. As in the evaluation studies of the original ToBI, a pairwise analysis was used to evaluate the consistency of the transcribers: the label of each transcriber was compared against the labels of every other transcriber for the particular aspect of the utterance. The 574 words were transcribed by the five subjects; thus a total of 5740 (574 x10 pairs of transcribers) transcriber-pair-words were produced. The following consistency rates were obtained: presence of pitch accent (73%), choice of pitch accent (69%), presence of phrase accent (82%), presence of boundary tone (89%), choice of phrase accent (78%), choice of boundary tone (85%), choice of break index (68%).

The level of consistency achieved for 4-Tone EVo transcription was somewhat lower than that reported for the original ToBI system. However, the differences in the agreement levels seem quite insignificant bearing in mind that 4-Tone EVo uses four tones instead of two! More extensive evaluation studies are currently underway to investigate the applicability of 4-Tone EVo in genuinely spontaneous discourse.

The *Laina* utterance might get the following 4-Tone EVo transcription, showing the creaky (non-modal) phonation increasing towards final position:

	Laina	*lainaa*	*Lainalle*	*lainan*.
%L	L+H*	L+H*	L+H*	l+H* l- l%

If, as above, we argue that the tune is best formalised as the tritonal sequence LHL, rather than as a sequence of two tones, and that both the initial L and the H are anchored to the initial, stressed syllable (i.e. starred tones are not needed), the *Laina* utterance would get the following transcription:

	Laina	*lainaa*	*Lainalle*	*lainan.*
%L	LHL	LHL	LHL	lHl l-l%

A similar utterance with words containing one-moraic first syllables would get the following type of pattern:

	Late	*lanaa*	*Lululle*	*lumet.*
%L	LHL	LHL	LHL	lHl l-l%

The system thus indicates L and H pitch targets but also shows if the phonation is in a non-modal mode (l and h) — note that l and h, as well as L and H, are relative, not absolute, pitch targets.

10.6 Intonation range

Above, we have discussed intonation patterns at sentence/utterance level, with a special reference to linguistic and attitudinal functions. It must be remembered that prosody is ever-present in speech also globally in that long-term average features of F_0 behaviour relate to the overall "liveliness impression" of a speaker's speech. To report the average F_0 value for speakers (or groups of speakers) is relatively straightforward, but the description of pitch range or intonation range is more problematic.

The first (and least recommended) approach is to describe pitch range with the linear Hertz scale. The problem is that this scale fails to make an appropriate normalisation for the non-linearity of pitch perception: a large change in frequency at the higher absolute pitch range is needed to produce the same perceptual effect as a smaller change at the lower absolute pitch range. Thus, with the linear scale, comparisons of speaker sex in pitch range are almost pointless. The second option is to convert the Hertz values into semitone values; the logarithmic semitone scale is closer to the human perceptual scale than the Hertz scale. The semitone scale has been extensively, although somewhat unsystematically, used in investigations of Finnish pitch range. The results have been somewhat inconsistent (over and above the obvious fact that intonation

range varies across speech situations and contexts), indicating either very narrow or very wide intonation ranges (see e.g. Lehessaari 1996 and Toivanen 2001, for detailed reviews of these studies). The descriptive problem seems to be that even the semitone scale is not completely appropriate from a view point of perception (the semitone scale actually retains some of the drawbacks of the linear scale). The third (and evidently the best) strategy is to use ERB measurements (Equivalent Rectangular Bandwidth). The ERB scale is based on the frequency selectivity of the human auditory system, and the scale is perceptually more appropriate for prosody than either the linear Hertz scale or the logarithmic semitone scale (Hermes and van Gestel 1991).

Toivanen (2001) investigated the prosody of Finnish English L2 speech in an experimental setting with native English speech as baseline data. Two groups of speakers (Finnish university students of English and native speakers of British English) read out a set of short standard texts, and the recorded speech data were analysed acoustically. Pitch range was described with the semitone scale and the ERB scale; the linear scale was used in some preliminary comparisons. A number of unsystematic differences in pitch variation between the two groups were found with the linear scale, while the semitone scale produced more consistent differences. The most systematic differences throughout the data, however, were detected using ERB measurements. The EBR scale enabled the conclusion that pitch variation in Finnish English L2 speech is indeed generally more limited than in native English speech. Clearly, the type of scale used for pitch analysis is critical, and it is recommended that in comparative cross-linguistic investigations of prosody, the perceptually relevant ERB scale should be considered as a first choice. To date, a large-scale, systematic analysis of the pitch range of colloquial Finnish utilising the ERB scale is yet to be carried out; until then, claims that Finnish voice range is “narrow“ or “monotonous” in comparison with other languages are not particularly convincing.

10.7 Intonation of Finnish as a second language

There are nowadays more and more speakers of Finnish as a foreign/second language. Russian, Somali and English speakers of Finnish are currently some of the biggest groups although the socio-economic backgrounds of these immigrant groups are highly divergent. Since 1990, Finland has received thousands of

Somalis fleeing civil war, thousands of Kurds from the Middle East, and thousands of refugees fleeing the Balkan conflicts.

Very little is known about the prosodic characteristics of these varieties of Finnish. It is plausible, however, that the speakers' native tongue exerts a highly important influence also on the prosodic features of second language. Below, we briefly report the results of a preliminary study (Toivanen, in preparation).

For the purposes of this investigation, Finnish speech data produced in the context of counselling discussions between Finnish student tutors (females in their twenties) and British students (females in their twenties) were used. The Finns were third-year AMK (university of applied sciences) students and the Britons were first-year AMK students. The Britons were persons who had moved to Finland at the age of 15–16; thus they had been exposed to Finnish for an average of 8–9 years, and they could be considered fluent (or semi-fluent). Each Briton had dual British and Finnish nationality. The speech material was produced by three Finnish students and ten British Finnish students, and the second language speech data was chosen for study.

The audio recordings were made in an anechoic chamber using a high quality equipment (48 kHz, 16-bit). The acoustic analysis was carried out with *f0Tool* (below, we will describe this algorithm in some detail). A written permission was obtained from each speaker. To ensure that no sensitive information would be available to the third party, any student could freely delete any speech material segment (of any length) from the digital tapes. The total duration of the English Finnish speech material (after the edition procedure described above) was 56 minutes (including pauses, e.g. hesitation breaks, naturally occurring in communication).

The recorded English Finnish speech data were analysed both auditively and acoustically. The speech data were divided into tone groups using acoustic and syntactic criteria (i.e. criteria based on pause occurrence and phrase and clause structure), and a tone choice was determined for each such tone group. An utterance or speaking turn (*lausuma* or *puheke* — see section 10.2.) containing several nuclear prominences was analysed as containing several tone groups. A pause was usually an important criterion for the demarcation of tone group boundaries (again, see section 10.2.). Note that in the analysis intonation was analysed in terms of (utterance-final) nuclear tones instead of tonal sequences (in any position in the utterance), so the possibility of a prevalence of LH sequences in the LHL tonal sequences was not revealed. That is, by investigating tonal

sequences throughout the utterances, rises might have been even more common (from this viewpoint, it could, of course, be argued that rises are common in first language Finnish speech as the LHL sequences, by definition, also contain rises). Table 7 shows the results of the present analysis.

Table 7. Distribution of tone choices (n=1594) in the non-native speech data: a nuclear tone classification.

Tone choice	Number	Percentage
low-fall	142	8.9%
high-fall	824	51.7%
rise-fall	63	4.0%
low-rise	129	8.1%
high-rise	112	7.0%
fall-rise	256	16.1%
level tone	48	3.0%
non-dynamic tone	20	1.3%

Although there are not any directly comparable statistical studies of tone choice in first language Finnish, it can be assumed, also on the basis of the above discussion of Finnish intonation, that falling tones will be overwhelmingly dominant in a Finnish speech context. One might estimate that the prevalence of falling tones in first language Finnish is clearly over 90% (and that rising tones maximally constitute, say, 5% of tones). In the second language Finnish speech data, by contrast, the prevalence of rising tones exceeded 30%. One can easily attribute this effect to the influence of the speaker's mother tongue, British English, and a more detailed investigation indeed reveals that the speakers systematically used rising tones to convey attitudinal meanings typically associated with English intonation.

The non-native Finnish speech data were also analysed in terms of voice quality features: for each intonation-group, the principal voice quality attribute was chosen auditively. "Modal" voice refers here to the most neutral type of phonation, with an absence of creak, falsetto, etc. An additional attribute (tense, lax, creak) was chosen if even one syllable of the intonation-group carried such a voice quality feature. Table 8 presents the most central results; the principal voice

quality characteristic is shown separately for each intonation-group type and thus also for each tone choice (since an intonation-group contains, by definition, one principal intonation contour).

Table 8. Voice quality in tone choices (n=1594) in the non-native speech data.

Tone choice	Number	Percentage
fall (modal)	862	54.1%
fall (modal, tense)	5	0.3%
fall (modal, lax)	60	3.8%
fall (creak)	39	2.4%
rise-fall (modal)	59	3.7%
rise-fall (modal, tense)	4	0.3%
rise (modal)	206	12.9%
rise (modal, tense)	11	0.7%
rise (modal, lax)	24	1.5%
fall-rise (modal)	174	10.9%
fall-rise (modal, lax)	82	5.1%
level tone (modal)	48	3.0%
other tone (modal)	2	0.1%
other tone (creak)	18	1.1%

It can be seen that intonation-groups with creak were very rare (3–4%). Again, we do not have directly comparable data on first language Finnish but it can be estimated that, for most native speakers, intonation-groups with creak (in final position in declaratives) would be very common.

An analysis of the interrelationship between intonation and discourse structure in the English Finnish data is currently underway. The preliminary investigation suggests that the speakers typically used the fall-rise on discourse markers (e.g. *joo*, *juu*, i.e. "yeah", and *ok*) to articulate epistemic stance or speaker attitude (position, standpoint). The fall-rise generally functioned as a face-threat mitigator, a frame or a delay device. Uncertainty or reservation were possible ingredients of the meaning, and the tone could also be viewed as

conveying a conventional implicature with specific content, i.e. a reservation, hesitation, etc. All these functions of the fall-rise are, of course, observable in native English intonation, as well. Interestingly, spoken English Finnish seems to adopt these first language intonation features quite facilely.

Indeed, intonational borrowing from English may be happening in Finnish more generally. In spoken Finnish, *ok*, uttered with a fall-rise, is now relatively common (at least in the speech of teenagers and young adults), having a back-channel function with a hint of reservation or even disbelief. Incidentally, *okra*, an intensified variant of *ok* used by adolescents in the Helsinki area, is, it seems, always uttered with a falling tone.

One may wonder in this context whether second language Finnish will eventually influence first language Finnish also prosodically. Could the result be attrition, i.e. the phenomenon of a first language being influenced by its second language varieties? Will rising intonations become more common in Finnish because non-native speakers of Finnish are not shy of using them? Will Finns become less persistent creakers when they learn that speakers with other first language backgrounds are not particularly fond of this voice quality feature? Recall from above (sections 3.2, 6.2.1 and 6.2.2) that foreign influence (mainly from English now) is strong on the consonant phoneme system and on word-initial and word-medial consonant sequences.

10.8 Emotional Finnish speech: research questions, data bases and tools

As has been discussed in the previous sections, there is the persistent stereotype that Finns do not utilise intonational/prosodic signals in speech as freely and intensively as speakers of some other languages (e.g. Italians). Stereotypically, it has been assumed that Finns tolerate long silences in conversation and are reluctant to engage in spontaneous small talk with strangers in communicative situations (again, consider the portrait of a Finn in a Kaurismäki movie). While there is empirical evidence on the syntactic prosodic aspects of spoken Finnish (some of which was presented above), the literature on the affective prosody is limited: the available empirical evidence concerning the purely affective aspects of spoken Finnish is fragmentary. There has been very little research on the vocal parameters of affective content in *continuous* spoken Finnish. For example, Laukkanen, Vilkman, Alku and Oksanen (1996, 1997) focus on very short units

(nonsense syllables with Finnish phonotactic structure) in their investigation of the vocal expression of emotion.

Here we present results of a recent research project on the vocal parameters of emotions in continuous spoken Finnish; to our knowledge, this is the first systematic study of the prosodic features of emotions in connected Finnish speech. The data were analysed from the viewpoint of the perception of emotions (i.e. from the viewpoint of listeners analysing the emotional content and utilising prosodic cues to aid the inference process). In addition, the data were utilised with the aim of the automatic classification/discrimination of emotions from speech: a statistical classifier was developed that uses prosodic cues to classify the emotions into predetermined categories. Thus we looked at the way in which emotions are expressed in continuous Finnish speech, and how human listeners and the computer can classify the emotions with the help of a number of acoustic/prosodic cues. We also present the speech corpus on which the research was based: in classification experiments, it is always necessary to collect a representative database, preferably as large as possible, and to systematically base the experiments on the corpus.

The *MediaTeam Emotional Speech Corpus* is currently the largest existing emotional speech corpus for continuous Finnish speech (Seppänen, Toivanen and Väyrynen, 2003). The corpus is used to investigate in detail the phonetic and phonological correlates of basic emotions in Finnish, and the results are used in developing speech corpus search engines (Toivanen and Seppänen, 2002). The speech material was produced by fourteen professional actors (eight men, six women) from Oulu City Theatre in Finland. The subjects were aged between 26 and 50, and all were speakers of the same northern variety of Finnish. The speakers simulated the following basic emotions while reading out a phonetically rich text of 120 words adapted from a newspaper article: neutral, sadness, anger, and happiness. The speakers were allowed to repeat the reading if they were not satisfied with the first rendition. Semantically, the text was as neutral as possible, describing features of a berry that grows in the northern parts of Finland. In addition to the monologue text, the speakers acted out two pre-written dialogues containing specific emotional lines of varying length. Thus the corpus contained linguistic units with specific emotional content ranging from short exclamations to monologues of approximately one minute in length.

The audio recordings were made in an anechoic chamber using high quality equipment, and the acoustic analysis was carried out with *f0Tool* (developed by

MediaTeam Language and Audio Technology Group). *f0Tool* is a speech analysis software implemented in the MATLAB language; *f0Tool* is a cepstrum-based voiced/unvoiced segmentation and time domain F_0 extraction algorithm using waveform-matching. The performance of the tool was tested with challenging speech material from radio conversations involving Finnish fighter pilots, and the accuracy of the tool was found to be quite comparable to the performance level of the existing standard speech analysis algorithms. The tool and the verification of its performance level are described in detail in Toivanen, Väyrynen and Seppänen (2004).

Currently, *f0Tool* is capable of analysing over 40 acoustic/prosodic parameters fully automatically from a speech sample of basically any length; the input required by *f0Tool* is an audio waveform file (Toivanen *et al*., 2004). The parameters are F_0-related, intensity-related, temporal and spectral features. Note that the term "segment" here refers to a part of the signal of varying duration, which may be realised as silence or as voiced or voiceless speech (i.e., the term does not designate any phonological unit here).

The general F_0-based parameters were e.g. the following: mean F_0, median F_0, maximum F_0, minimum F_0, F_0 range, 5th fractile of F_0 and 95th fractile of F_0. The parameters describing the dynamics of F_0 were e.g. the following: average F_0 fall/rise during a continuous voiced segment, average steepness of F_0 fall/rise, maximum F_0 fall/rise during a continuous voiced segment, and maximum steepness of F_0 fall/rise. The intensity-related parameters were e.g. the following: mean RMS intensity, median RMS intensity, intensity range, 5th fractile of intensity, 95th fractile of intensity, and the range between the fractiles. The temporal parameters were e.g. the following: average duration of voiced segments, average duration of unvoiced segments shorter than 300 ms, maximum duration of voiced segments, and maximum duration of silence segments. Ratio parameters were e.g. the following: ratio of speech to long unvoiced segments and ratio of silence/speech segments. The spectral features concerned the proportion of low-frequency energy (below 500/1000 Hz). Additional parameters were jitter and shimmer. All these parameters are common phonetic characteristics of voice quality. This is not the place to define the parameters in detail; below, we describe some central parameters briefly.

10.9 Emotion in spoken Finnish: evidence from classification experiments

Speaker-independent classification was performed using the k-Nearest-Neighbour classifier (kNN), which is applied as a standard non-parametric method in statistical pattern recognition; leave-one-out was used for evaluating classifier performance. The level of automatic classification of emotions reached a level of just below 70% with the prosodic parameters given in Table 9 that represent seven dimensions in the classification procedure, intensity range being the single most important cue. Note that intensity range alone produced a classification capacity of over 51.1%, and that intensity range and maximum F_0 rise during a voiced segment together yielded a classification rate of 54.6%, etc.

Table 9. Emotional cues in spoken Finnish for the computer.

Acoustic feature	Cumulative classification accuracy
intensity range	51.1%
maximum F_0 rise during a voiced segment	54.6%
ratio of silence-to-speech	63.2%
5%–95% F_0 range	65.0%
shimmer	66.1%
jitter	68.6%
intensity variation	69.6%

The highest F_0 value and the lowest F_0 value are absolute values, and are not often very useful parameters as they may be, in effect, "accidental" values because they often represent (unintentional) shifts into the falsetto and creak registers, respectively. Therefore, a more useful technique is to compute the 5^{th} percentile of F_0 (instead of the absolutely lowest F_0 value) and the 95^{th} percentile of F_0 (instead of the absolutely highest F_0 value). The total F_0 range is the absolute difference between the highest observed F_0 value and the lowest observed value. This is one way of describing the dynamics of F_0 variation in the sample but, again, a better measure is the one based on percentiles (e.g. the *5^{th}–95^{th} percentile range*), as in Table 9.

Jitter can be defined as the amount of random cycle-to-cycle variation between adjacent pitch periods in vocal fold vibration; it is thus a measure of F_0 perturbation. *Shimmer* is the amount of cycle-to-cycle variation in amplitude between adjacent pitch periods.

Human classification experiments were performed in the form of listening tests. The listeners were students in a junior high school, aged between 14 and 15: fifty-one subjects (27 males, 24 females) participated as volunteers. All listeners were speakers of the same northern variety of Finnish as the actors. The emotional labels to choose between were limited to the intended emotions; distracters were not used. The average emotion discrimination performance of the listeners was 77%. The exact performance levels of the classification and the full list of the best parameters for the first data set can be found in Toivanen *et al*. (2004).

The existing literature suggests that the computer achieved quite a good discrimination rate. It has been argued that, in a speaker-independent classification task, as in our experiment, the performance level can reach 60–70% for three basic emotions (ten Bosch 2003). Looking at the best feature vector in the classification task, it was observed that, to express emotion vocally, the speakers used cues largely similar to those reported for other languages, i.e. variations in energy, speech rate and pitch. The optimal set of parameters in the classification procedure consisted of intensity range, maximum F_0 rise during a continuous voiced segment, ratio of silence-to-speech, 5%–95% F_0 range, shimmer, jitter, and intensity variation. This set clearly reflects the "liveliness" of the speech: intensity range, F_0 range, the dynamics of F_0 change as well as the amount of speech within a speaking turn obviously correlate with the activity level of the speech situation and the speaker. It can thus be concluded that Finns use prosody to express affect in speech in a way that must be essentially similar to the vocal expression of emotion reported for major languages such as English and French. Showing that the same prosodic parameters are utilised in the emotion portrayals through voice, and demonstrating that emotional spoken Finnish is not qualitatively different from other languages, our research finding hopefully serves to dispel some myths about the characteristics of Finnish speech.

An interesting product of our experiment is the 7% difference between the performance levels for the computer and the human listeners, demonstrating that the human listeners utilised acoustic/prosodic parameters unavailable to the computer. The computer can utilise only automatically computable prosodic

primitives, while the human listener also pays attention to the linguistically relevant prosodic phenomena.

As was pointed out above, in spoken Finnish, the basic non-affective utterance contains a descending F_0 curve with rising-falling peaks in the syllables of the accentuated words. Our point here is that accents (i.e. rhematic, contrastive and emphatic accents) which are signalled tonally probably tend to co-occur with special emotional content in speech. The human listener will hear these accents as discrete phonological phenomena, but the classifier (i.e. the computer) is not capable of this — in its current form. Thus the human listener has access to more information than the computer in evaluating the affective dimensions of speech — as the observed performance level difference in our data suggests. In emotional Finnish, the dynamic aspects of F_0 variation (e.g. maximum F_0 rise) thus probably have an important role from the perceptual viewpoint. In addition to signalling the beginning of accent, F_0 rises in all likelihood also occur in utterances which are "globally emotional": they do not just mark off single accentuated words with contrastive and/or emphatic content, but they represent speaking turns which are emotional throughout. As was pointed out above, a rising intonation is (still) relatively rare in SSF unless an emotional dimension is intended. Utterances with high-rising tones can be assumed to convey strong emotional meanings (annoyance, incredulity, etc.) in spoken Finnish. Again, it must be noted that the current classifier does not "hear" these syntactic features of rising F_0 movements (in final position) in an utterance. By contrast, the human listeners can be expected to be fully aware of this kind of "marked" prosody in a speaking turn. It should also be noted that these phonological (emotion-related) F_0 features certainly exist in spoken Finnish regardless of the possibility that Finnish is not, phonologically, as tonal as some other languages. The degree of tonality may be "small" only in comparison with other languages: the language-specific tonal features are quite distinct in Finnish to separate contrastive accents from thematic ones, and emotional speech from non-emotional speech.

The results of the classification experiments offer (indirect) support for the hypothesis that discrete non-gradable phonological features — accents and utterance-level intonation contours — also convey affective content in Finnish. This has implications for the development of classification methods. It will not be enough to concentrate on the automatically measurable phonetic variables; at some point, the classifier must tackle the more abstract prosodic patterns if the aim is to ultimately improve the emotion discrimination performance level. An

important future direction in the development of classification methods would be to model the abstract F_0 phenomena in a computable (i.e. computer-interpretable) way. There is no reason to assume that this would be an impossible task in the long run. Essentially, what is needed is the gradual development of language-specific models of legitimate phonological F_0 features, which the classifier must be trained to recognise. Eventually, in the classification procedure, the constantly varying prosodic features and the more abstract features must be combined.

On the basis of the above, some conclusions about the cues for emotion in spoken Finnish seem possible. First, features of F_0 and intensity have been found to accompany emotional Finnish speech — this is probably a universal phenomenon in the expression of emotion. Second, the performance level of the human emotion classification exceeds that of the automatic classification. Although this is not surprising in itself, we suggest that phonological features of F_0 variation, especially rising F_0, are emotion-carrying features in spoken Finnish, in addition to the global constantly varying average features of F_0, intensity, duration, etc. Also in this respect, it can be argued that Finnish, a small language in a small language group, is not qualitatively different from major Indo-European languages. This finding contradicts stereotypical notions of (the lack of) emotionality in the Finnish language. In languages in general, prosodic parameters are hierarchically organised as concrete ("phonetic" or "paralinguistic") and as more abstract ("phonological" or "linguistic") phenomena, and there is no reason to assume that some of these levels would be irrelevant from the viewpoint of the vocal communication of emotion. Finally, the results suggest that contrastive research on human vs. computer categorisation of emotions is promising, and that, in the near future, computer recognition of human vocal emotions may approach a natural state, yielding access to new exciting product applications.

10.10 A summary of some things glottal

In this section we summarise some of the uses of the glottis in Finnish; most of these uses have already been mentioned earlier in this book. We do not summarise the segmental and prosodic uses of normal modal phonation (such as the difference between voiced and voiceless segments, or F_0 variations signalling accent and intonation), but other, cross-linguistically perhaps less common uses

of the glottis, such as non-modal phonation, whisper, and other sounds originating in the glottis.

Finnish has three glottal consonants, [h], [ɦ] and [ʔ]. The first two occur as allophones of /h/, but the glottal mechanisms of these consonants have also aphonematic uses, and the glottal stop has only aphonematic uses. Suomi (to appear) observed that an epenthetic [h] was regularly inserted to the end of utterances consisting of a single monomoraic word, e.g. an utterance consisting of the word *se* 'it' was pronounced [seh]. But such a segment was not inserted to the end of utterances consisting of a single dimoraic word, such as *joo*, except quite sporadically. It seems then that the function of this epenthetic [h] is to guarantee that even the minimal utterance consists of at least one segment after the only mora, a segment that was interpreted as a phonetic mora. The [h]-like noise observed in Suomi (2007) in contrastively accented monomoraic words like *se*, a long distance from utterance end, seems to have a similar function: to act as an aphonematic filler that signals that the word is accented. Utterance-medially a pause is available as an alternative filler, but not utterance-finally (as in the minimal utterances). In both contexts, to avoid confusion with a double vowel, the modally voiced vowel portion cannot be lengthened.

Then there are the [ɦ]-like, [h]-like and whispery endings observed at the ends of non-minimal utterances. Lehtonen (1970: 45) noted that in his informants' speech the final part before utterance end was often a completely voiceless whisper and added, in a footnote, that "[t]he voicelessness of the 'tail' of an utterance is not a consequence of an extraordinarily careless speech habit but a very common feature in spoken Finnish". In these noises the moraic structure of the utterance-final word or syllable does not seem to have any effect on the duration or probability of occurrence of the final voiceless portion: in Myers & Hansen (2006) the mean durations of the voiceless portions were the same after single vowels (69 ms) and after double vowels (71 ms), and the were apparently equally common in both contexts. Before modal voice turns to voicelessness, often a period of breathy voice (similar to that in [ɦ]) intervenes.

It is not clear what the relative frequencies of breathiness, creak and whisper are before utterance boundaries; all these seem to be more common in declarative statements than in questions. Ogden (2001) analysed spontaneous conversations and reported that final creaky voice, followed by whisper and exhalation was used to mark the end of conversational turn. Ogden observed more instances of creaky voice than did Nakai *et al*. (to appear) and Myers & Hansen (2006) who report

that the speakers "generally had a period of voicelessness at the end of the final vowel, in which the formants of the final vowel were continued in aperiodic noise, as in an [h]" (p. 170–171). This difference may be due to a difference in the speaking situations: Ogden studied the ends of conversational turns, the other authors studied the ends of non-conversational utterances (and Lehtonen's comment above also concerns non-conversational utterances). It is possible that creaky voice and glottal stops have more conversational functions than voiceless speech ([h]-like or whispery endings). Bruce (1998: 133) suggests that creak at phrase end is common in Swedish whereas breathy phonation seems to more usual in Finnish. For all we know this may be true.

The glottal stop never occurs word-internally in Finnish (in non-corrupt pronunciations), but it does occur elsewhere: between x-morphemes and vowel-initial words (with the reservations mentioned in section 5.2); before phonologically vowel-initial words, in dialect interviews at least, in many dialects as reported by Itkonen (1964); in colloquial conversational speech, as reported by Lennes *et al*. (2006); and in single-word colloquial replies consisting of the reduced forms of *en* 'I don't' and *on* 'it is' (pronounced as [ʔeh] ja [ʔoh]), as reported by Suomi (to appear). The last two observations may be instances of the same conversational occurrence of glottalisation.

Ogden (2001) found final creaky voice, followed by whisper and exhalation, to mark the end of conversational turn. Lennes *et al*. (2006) in turn concluded that glottal stops with complete closure can be used, apart from other purposes, also for signalling one's intention to continue speaking, i.e. for holding the turn. Scanty as such observations are, they suggest a division labour.

Creaky phonation was also mentioned (in Chapter 2) as an increasingly common hallmark of young, especially female speakers, a property that is not limited to prosodic boundaries but characterises speech as a whole. Non-modal voice qualities are often prosodic characteristics of emotional discourse (sadness, anger, etc.). In the way emotions are signalled in speech, Finnish does not seem to be qualitatively different from major Indo-European languages.

To conclude this summary section, noises and silences like the consonants [h], [ɦ] and [ʔ] are used for many aphonematic purposes in Finnish, purposes in which the noises and silences do not constitute parts of words. Other consonants do not serve similar purposes: in boundary lengthening, also other consonants are lengthened, but the consonants are part of the post-boundary word, as in *Mene pois!* [menep:ois] 'Go away!' in which the second word, in any reasonable

interpretation, begins with /p/. Indirectly at least, the aphonematic uses of [h], [ɦ] and [ʔ] may be argued to motivate the postulation of glottals as a major class of consonants in addition to obstruents and resonants.

11 Sound structure and orthography

The relationship between the sound structure and orthography in Finnish may deserve some comments. Firstly because, as has been discussed above, orthography may have had a role to play in how foreign languages have influenced the recent developments of the Finnish sound pattern, notably the emergence of new consonant phonemes. Secondly, because the relationship between sound structure and orthography is in many respects different from that in many other languages with an alphabetical writing system, and because it is often mentioned as a distinguishing characteristic of Finnish — not only by laymen, but also in descriptions by professional linguists — that the language "is written (exactly) as it is spoken" (or *vice versa*). Such claims of course overlook the fact that orthography never shows allophonic details, and that orthography almost completely overlooks utterance prosody; punctuation marks may at best give a very crude representation of this. Nevertheless, claims of how Finnish is spoken vis-à-vis the standard orthography do capture something about the Finnish orthography in comparison to those of many other languages, but at the same time they must be taken with a grain of salt, even if only phonemic structure of speech is considered. On the whole, with some reservations mentioned below, standard Finnish orthography follows the basic alphabetical principle: there is, by and large, a one-to-one correspondence between phonemes and graphemes such that a given phoneme is represented by a given grapheme. There is only one fully systematic exception in the native vocabulary that pertains to all varieties of spoken Finnish that have /ŋ/ as a phoneme: Because the Latin alphabet has no letter corresponding to the phoneme /ŋ/, and because no new letter has been adopted for this purpose, /ŋ/ is represented by <n> (before <k> (as in *kenkä* 'shoe'), before <g> (as in *Englanti* in which no /ɡ/ is pronounced), after <g> (as in *kognitio* in which the sequence /ŋn/ is pronounced), and /ŋŋ/ is represented by <ng> (as in *kengän* 'of shoe' and in *tango*).

As already mentioned in Chapter 1, Standard Spoken Finnish (SSF) was based on Standard Written Finnish (SWF). The standard orthography, like SSF, is taught at schools and is used in the written media and in most of the literature (excluding literature specifically written in a dialect). Today, if a speaker has the maximum paradigm of 17 consonant phonemes, and if the speaker's pronunciation, in a given situation, follows SWF very closely, using only the full forms of words, then it can be said that there is a one-to-one correspondence

between the phonemes of speech and the graphemes of the standard orthography (excluding /ŋ/ and /ŋŋ/). However, such a manner of speaking is becoming increasingly rare even in formal speaking situations. Instead, speakers very often, even if they speak a variety of SSF and not their local dialect, use reduced forms of many words; for example, deletions of word-final vowels in inflectional endings and of word-internal /d/'s are common. And speakers who do not have the maximum paradigm of 17 consonant phonemes may pronounce e.g. the words *paletti* 'palette' and *baletti* 'ballet' alike. Consequently, a speaker may very well utter a colloquial sentence about the ballet that can be phonemically represented as /kyl must̪ pɑlet̪ist̪ pit̪æs puhuu/ 'in my opinion, the ballet should be talked about'. Yet, if the sentence were quoted in the press, it would most probably be rendered in print as <Kyllä minusta baletista pitäisi puhua>. In this printed rendition, which follows the norms of SWF, there are at least seven graphemes (those underlined) that have no corresponding phoneme segment in the quoted utterance, namely <ä> at the end of the first word, <in> medially and <a> at the end of the second word, the final vowel in the third and fourth word, and <i> in the middle syllable of the fourth word). In addition, there are two graphemes in the printed rendition that represent a "wrong" phoneme: the initial consonant in the third word (/p/ was pronounced, but it was written <b>), and the last vowel of the last word (it is written <a>, although /u/ was pronounced). Thus, in this invented but quite possible example, the standard orthography contains seven graphemes that have no phonetic material in the utterance to motivate them (i.e., there are seven "silent letters"), and two graphemes that violate the principle that a given grapheme always represents the same phoneme. In this sense, then, Finnish is by no means always written as it is spoken: more exactly, the phonemic sound structure is not always reflected by the graphemes in writing in a one-to-one correspondence. At the same time, a speaker of Finnish usually knows how to spell a novel word heard as pronounced in its full form, given that the word contains no phonetic segments that are ambiguous in the speaker's phoneme system. The spelling conventions of SWF have been established rather recently, the pronunciation of SSF has not changed very much since then, but it is not unlikely that the discrepancy between the phonemes of SSF and the graphemes of SWF will become larger in the future, unless the spelling conventions are changed to observe changes in pronunciation.

The graphemes <v> and <w> are non-distinct in Finnish in the sense that the choice between them implies no difference in pronunciation (and in telephone

catalogues, for example, it is the next grapheme that determines the order: the name Varis is listed before Wellin). In the olden times, <w> was used in writing down ordinary native words and there was much variation in the use of the two graphemes among authors; today <w> is not used in writing down native words in the standard orthography, it is used only in trade names and the like to achieve a particular effect, e.g. in commercial logos attempting to arouse nostalgic feelings (and open the customers' purses). Nevertheless, it is not unusual to see the name of a famous university written as Harward.

Today, the correspondence between phonemes and graphemes is more one-to-one in Finnish than it is in e.g. English (in which there are, due to the Great Vowel Shift and other changes, many "silent letters" that are never pronounced, many words pronounced identically but written differently etc.). One consequence of this state of affairs is that it is possible in Finnish, to a considerable extent, to represent in writing differences between dialects (but not between local varieties of SSF, as the differences usually only concern prosodic properties which orthography ignores). For example, the Finnish word for English *terrible* can be written as e.g. <kauhea> (the normative orthographic form), as <kauhee>, <kauhia>, <kauhi> or <kaahee>, and these different orthographic forms, reflecting clear phonemic differences in pronunciation among varieties, are very much able to convey information of the variety of Finnish in question. Similarly e.g. the (nominative form of) first person singular pronoun can be written at least as <minä> (the conventional orthographic form), <mnää>, <mää>, <mä> and <mie> to reflect differences between varieties (and similarly <sinä>, <snää>, <sää>, <sä> and <sie> for 'thou'). Not surprisingly, dialect books and poems are a popular literary genre.

It is difficult to see how corresponding differences could be represented in writing systems like that of English, in which letters have a much more remote relationship to pronunciation. In sum, because Finnish has been written down for a relatively short time, because the normative SWF orthography has been established rather recently, and because, consequently, the phoneme-grapheme correspondence is still a relatively close one, Finnish orthography is able to reflect more details of the sound structure of the spoken language than is the case in languages in which the official writing conventions have been established much earlier, and in which pronunciation has had more time to change after the establishment of the normative writing conventions.

References

Ahonen-Mäkelä, R (1975) *Ilo puhua. Ajatuksia puheilmaisusta*. The Finnish Literature Society, Helsinki.

Arvaniti, A, Ladd DR & Mennen I (1998) Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.

Beckman, ME & Ayers GM (1993) *Guidelines for ToBI Labelling*. Linguistics Department, Ohio State University.

Bolinger D (1972) Accent is predictable (if you're a mind reader). *Language* 48: 345–359.

Bolinger D (1989) *Intonation and its Uses*. Edward Arnold, London.

Bosch L ten (2003) Emotions, speech and the ASR framework. *Speech Communication* 40: 213–225.

Brazil D, Coulthard M & Johns C (1980) *Discourse Intonation and Language Teaching*. Longman, London.

Bruce G (1977) *Swedish word accents in a sentence perspective*. Gleerup, Lund.

Bruce G (1998) Allmän och svensk prosodi. Praktisk lingvistik 16. Institutionen för lingvistik, Lunds universitet.

Cruttenden A (1997) *Intonation* (2nd edition). Cambridge University Press, Cambridge.

Garlén C (1988) *Svenskans fonologi*. Studentlitteratur, Lund.

Gordon M (1997) Phonetic correlates of stress and the prosodic hierarchy in Estonian. In: Ross J & Lehiste I (eds) *Estonian prosody: Papers from a symposium*. Institute of Estonian language, Tallinn: 100–124.

Häkkinen K (1978) Eräistä suomen kielen äännerakenteen luonteenomaisista piirteistä ja niiden taustasta. An unpublished licentiate thesis. Department of Finnish and General Linguistics, University of Turku.

Hakulinen L (1961) *The Structure and Development of the Finnish Language*. Translated by John Atkinson. Indiana University Publications, Uralic and Altaic Series, Vol. 3. Indiana University, Bloomington.

Hart J (1551) *The Opening of the Unreasonable Writing of Our Inglish Toung*. London.

Hart J (1569) *An Orthographie*. London.

Heringer J & Wolontis M (1972) Focus in Finnish. *Papers from the Eighth Regional Meeting of Chicago Linguistic Society* 8: 88–95.

Hermes DJ & van Gestel JC (1991) The frequency scale of speech intonation. *Journal of the Acoustical Society of America* 90 (1): 97–102.

Hirvonen P (1970) *Finnish and English Communicative Intonation*. Publications of the Phonetics Department of the University of Turku 8, Turku.

Hulst H van der (1985) Ambisyllabicity in Dutch. In: Bennis H & Beukema F (eds) *Linguistics in the Netherlands 1985*. Foris, Dordrecht: 57–66.

Huss V (1978) English word stress in the post-nuclear position. *Phonetica* 35: 86–105.

Iivonen A (1974) Äännekeston riippuvuus sanan pituudesta irrallaan äännetyissä sanoissa. *Virittäjä* 78: 134–151.
Iivonen A (1978) Is there interrogative intonation in Finnish?. In: Gårding E, Bruce G & Bannert R (eds) *Nordic Prosody: Papers from a Symposium*. Lund University Press, Lund: 43–53.
Iivonen A (1998) Intonation in Finnish. In: Hirst D & Di Cristo A (eds) *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, Cambridge: 311–327.
Itkonen T (1964) *Proto-Finnic final consonants: Their history in the Finnic languages with particular reference to the Finnish dialects. 1:1. Introduction. The history of -k in Finnish*. The Finnish Literature Society, Helsinki.
Järvikivi J, Aalto D, Aulanko R & Vainio M (2007). Perception of vowel length: Tonality cues categorization even in a quantity language. In Trouvain J & Barry WJ (eds) *Proceedings of the 16th International Congress of Phonetic Sciences*. Universität des Saarlandes, Saarbrücken, Germany: 693–696.
Karlsson F (1969) Suomen yleiskielen segmentaalifoneemien paradigma. *Virittäjä* 73: 351–362.
Karlsson F (1983) *Suomen kielen äänne- ja muotorakenne*. WSOY, Juva.
Karlsson F (1999) *Finnish. An Essential Grammar*. Routledge, London.
Karlsson F (2005) Phonotactic complexity of Finnish nouns. In: Arppe A, Carlson L, Lindén K, Piitulainen J, Suominen M, Vainio M, Westerlund H & Yli-Jyrä A (eds) *Inquiries into words, constraints and contexts. Festschrift in the honour of Kimmo Koskenniemi on his 60th birthday*. Gummerus Printing, Saarijärvi: 65–170.
Kaun A (2004) The typology of rounding harmony. In: Hayes B, Kirchner R & Steriade D (eds) *Phonetically Based Phonology*. Cambridge University Press, Cambridge: 87–117.
Kuronen M (2000) *Vokaluttalets akustik i sverigesvenska, finlandssvenska och finska*. Jyväskylä University Printing House, Jyväskylä.
Ladd DR, Faulkner D, Faulkner H & Schepman A (1999) Constant ‘segmental anchoring’ of F_0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–1554.
Ladefoged P & Maddieson I (1996) *The Sounds of the World’s Languages*. Blackwell, Oxford.
Laukkanen A-M, Vilkman E, Alku P & Oksanen H (1996) Physical variations related to stress and emotional state: a preliminary study. *Journal of Phonetics* 24: 313–335.
Laukkanen A-M, Vilkman E, Alku P & Oksanen H (1997) On the perception of emotions in speech: The role of voice quality. *Logopedics Phoniatrics Vocology* 22: 157–168.
Lehessaari AL (1996) *Alkoholin vaikutus puheen prosodiikkaan*. Publications of the Department of Phonetics 40. University of Helsinki.
Lehiste I (1972) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51: 2018–2024.
Lehtonen J (1970) *Aspects of Quantity in Standard Finnish*. Studia Philologica Jyväskyläensia VI, Jyväskylä.

Lehtonen J (1974) Sanan pituus ja äännekestot. *Virittäjä* 78: 152–160.

Lennes M, Aho E, Toivola M & Wahlberg K (2006) On the use of the glottal stop in Finnish conversational speech. In: Aulanko E, Wahlberg L & Vainio M (eds) *The Phonetics Symposium 2006*. Publications of the Department of Speech Sciences, University of Helsinki: 93–102.

Maddieson I (1984) *Patterns of Sounds*. Cambridge University Press, Cambridge.

Marjanen K (1932) Puheen emotionaalisesta sävelkulusta. Puhepedagogiikan psykologisen pohjan valaisua. *Virittäjä* 36: 279–293.

Monola T (1976) *Aino Kallaksen ja Marja-Liisa Vartion proosarytmin vertailua. Kolmen proosakatkelman rytmianalyysi.* Publications of the Department of Phonetics 28. University of Helsinki.

Mustanoja L & O'Dell M (2007) Suomen d ja r sosiofoneettisessa kentässä. *Virittäjä* 111: 56–67.

Myers S & Hansen B (2006) Neutralization of vowel length contrasts in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory* 25: 157–193.

Nakai S, Kunnari S, Turk A, Suomi K & Ylitalo R. Utterance-final lengthening and quantity in Northern Finnish. To appear in *Journal of Phonetics*.

O'Connor JD & Arnold GF (1973) *Intonation of Colloquial English*. Longman, London.

O'Dell M (2003*) Intrinsic Timing and Quantity in Finnish*. Acta Universitatis Tamperensis 979. Tampere University Press. URL: http://acta.uta.fi/teos.phtml?9540

O'Dell M & Nieminen T (2006) Tahdin ajoitus suomessa oskillaattorimallin näkökulmasta. In: Aulanko R, Wahlberg L & Vainio M (eds) *The Phonetics Symposium 2006*. Publications of the Department of Speech Sciences, University of Helsinki: 53: 134–143.

Ogden R (2001) Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association* 31: 139–152.

Peltonen V (1901) *Puhetaito*. WSOY, Porvoo.

Routarinne S (2003) *Tytöt äänessä: parenteesit ja nouseva sävelkulku kertojien vuorovaikutuskeinoina*. The Finnish Literature Society, Helsinki.

Sovijärvi A (1956) *Johdanto yleisen fonetiikan peruskurssiin*. University of Helsinki Press (handout).

Seppänen T, Toivanen J & Väyrynen E (2003) MediaTeam Speech Corpus: A first large Finnish emotional speech database. *Proceedings of the 15th International Congress of Phonetic Sciences* (Barcelona) 3: 2469–2472.

Sluijter A & van Heuven V (1996) Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471–2485.

Sulkala H & Karjalainen M (1992) *Finnish. Descriptive Grammars*. Routledge, London.

Suomi K (1980) *Voicing in English and Finnish stops*. Publications of the Department of Finnish and General Linguistics of the University of Turku 10, Turku.

Suomi K (1983) Palatal vowel harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* 6: 1–35.

Suomi K (1990) Huomioita yleiskielen konsonanttien yhdistelyrajoituksista ja pohjalaismurteiden epenteettisestä vokaalista. *Virittäjä* 94: 139–160.
Suomi K (1998) *Electropalatographic investigations of three Finnish coronal consonants*. Linguistica Uralica XXXIV: 252–257.
Suomi K (2005) Temporal conspiracies for a tonal end: segmental durations and accentual f_0 movement in a quantity language. *Journal of Phonetics* 33: 291–309.
Suomi K (2006) Suomen segmenttikestojen määräytymisestä. *Virittäjä* 110: 483–503.
Suomi K (2007) On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35: 40–55.
Suomi K (2008) Suomen kvantiteettiopposition fonologisesta tulkinnasta. *Virittäjä* 112: 403–408.
Suomi K. Minimaalisista ja muista lyhyistä puhunnoksista. To appear in *Virittäjä*.
Suomi K (Submitted) Durational elasticity for accentual purposes in Finnish.
Suomi K (In preparation) Segmental anchoring of accent and speaking rate in Finnish.
Suomi K., McQueen J & Cutler A (1997) Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36: 422–444.
Suomi K., Toivanen J & Ylitalo R (2003) Durational and tonal correlates of accent in Finnish. *Journal of Phonetics* 31: 113–138.
Suomi K & Ylitalo R (2004) On durational correlates of word stress in Finnish. *Journal of Phonetics* 32: 35–63.
Suomi K, Toivanen J & Ylitalo R (2006) *Fonetiikan ja suomen äänneopin perusteet.* Gaudeamus, Helsinki.
Terken J & Hermes D (2000) The perception of prosodic prominence. In: Horne M (ed) *Prosody: Theory and Experiment*. Kluwer, Dordrecht: 89–127.
Toivanen J (2001) *Perspectives of Intonation: English, Finnish and English Spoken by Finns*. Peter Lang Verlag der Wissenschaften, Frankfurt am Main.
Toivanen J (2006) Evaluation study of '4-Tone EVo': a multimodal transcription model for emotion in voice in spoken English. In: Toivanen J & Henrichsen PJ (eds) *Current Trends in Research on Spoken Language in the Nordic Countries*. Center for Computational Modelling of Language, Department of Computational Linguistics, Copenhagen Business School: 139–140.
Toivanen J (in preparation) Intonation of Finnish as a second language in student talk.
Toivanen J & Seppänen T (2002) Prosody-based search features in information retrieval. *Proceedings of FONETIK 2002* (Stockholm): 105–108.
Toivanen J, Seppänen T & Väyrynen E (2003) Automatic recognition of emotions in spoken Finnish: preliminary results and applications. *Proceedings of Prosodic Interfaces (IP2003)*: 85–89.
Toivanen J, Väyrynen E & Seppänen T (2004) Automatic discrimination of emotion from spoken Finnish. *Language and Speech* 47: 383–412.
Tuomi T (1972) *Reverse dictionary of Modern Standard Finnish*. The Finnish Literature Society, Helsinki.

Vainio M, Aalto D, Järvikivi J & Suni A (2006) Quantity and tone in Finnish lexically stressed syllables. In: Belotel-Grenie A & Grenie M (eds) *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages*: 121–124.

Välimaa-Blum R (1993) A pitch accent analysis of intonation in Finnish. *Ural-Altaische Jahrbucher* 12: 82–89.

Vihanta V (1978) *Les voyelles toniques du français et leur realisation et perception par les etudiants finnophones*. Studia Philologica Jyväskyläensia XII, Jyväskylä.

White L (2002) English speech timing: a domain and locus approach. University of Edinburgh PhD dissertation.
URL: http://www.cstr.ed.ac.uk/projects/eustace/dissertation.html

Wiik K (1965) *Finnish and English vowels*. Annales Universitatis Turkuensis, 94. Turku University, Turku.

Xu Y (2005) Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220–25.

Oulun yliopisto • Humanistinen tiedekunta
University of Oulu • Faculty of Humanities

ISBN 978-951-42-8983-5
ISSN 1796-4725

KANSI RAIMO AHONEN